Under the Radar:

The Effects of Monitoring Firms on Tax Compliance *

Miguel Almunia
University of Warwick

David Lopez-Rodriguez
Banco de España

March 14, 2017

Abstract

This paper analyzes the effects of size-dependent tax enforcement on firms' tax compliance. We exploit quasi-experimental variation generated by a Large Taxpayers Unit (LTU) in Spain, which monitors firms with more than 6 million euros in reported revenue. Firms strategically bunch below the eligibility threshold in order to avoid stricter tax enforcement. The response is stronger in sectors where transactions leave more paper trail, suggesting that monitoring effort and the traceability of information reported by firms are complements. We estimate that there would be substantial welfare gains from extending stricter tax monitoring to smaller businesses.

Keywords: tax enforcement, firms, bunching, Spain, Large Taxpayers Unit (LTU).

JEL codes: H26, H32.

*Almunia: m.almunia@warwick.ac.uk, University of Warwick Department of Economics and Centre for Competitive Advantage in the Global Economy (CAGE). Lopez-Rodriguez: david.lopezr@bde.es, Banco de España. A previous version of this paper circulated under the title *"Heterogeneous Responses to Effective Tax Enforcement: Evidence from Spanish Firms"*. We thank Emmanuel Saez, Alan Auerbach, Fred Finan and Ted Miguel for constant support and encouragement throughout this project. We gratefully acknowledge many useful comments and suggestions from Juan Pablo Atal, Pierre Bachas, Henrique Basso, Michael Best, David Card, Lorenzo Casaburi, Raj Chetty, Francisco de la Torre, François Gerard, Jonas Hjort, Simon Jäger, Anders Jensen, Juan F. Jimeno, Henrik Kleven, Rafael Frutos, Attila Lindner, Justin McCrary, Craig McIntosh, Adair Morse, Gautam Rao, Daniel Reck, Ana Rocca, Michel Serafinelli, Monica Singhal, Juan Carlos Suarez Serrato, Joel Slemrod, Victoria Vanasco, Andrea Weber, Danny Yagan, Owen Zidar and numerous seminar participants. Letizia Borgomeo and Gonzalo Gaete provided excellent research assistance. Almunia gratefully acknowledges financial support from Fundación Rafael del Pino and the Burch Center for Tax Policy and Public Finance. Any views expressed in this paper are only those of the authors and should not be attributed to the Banco de España.

1 Introduction

In recent decades, many tax administrations have established special units to monitor and improve the tax compliance of the largest taxpayers, following the advice of multilateral organizations (IMF, 2002; OECD, 2011). The expected benefits from establishing large taxpayers units (LTUs) include the ability to raise higher tax revenues and, more generally, increasing the efficiency of tax administration. Despite the widespread adoption of this size-dependent tax enforcement strategy around the world, there is very little micro-level evidence the effects of LTUs on firm behavior. Moreover, we are not aware of any existing empirical research that evaluates the welfare effects of size-dependent tax enforcement.

In this paper, we study the effects of size-dependent tax enforcement on firms' tax compliance and the welfare implications of this policy. In particular, we analyze how firms respond to the increased monitoring effort created by the eligibility cutoff of the Spanish Large Taxpayers Unit. This LTU focuses exclusively on firms with total operating revenue above €6 million, a threshold established in 1995 that has remained constant in nominal terms ever since. The LTU has more auditors per taxpayer than the rest of the tax authority, and these auditors are on average more experienced and better trained to deal with the most complex taxpayers. Thus, while firms above and below the eligibility threshold face the same tax schedule and essentially the same compliance procedures, monitoring effort changes discretely at this arbitrary cutoff. This allows us to study to what extent stricter monitoring interacts with the availability third-party information to increase firms' tax compliance. We first derive a simple theoretical model to analyze how firms are expected to respond to the increase in monitoring, and we then test these predictions exploiting the quasi-experimental variation generated by the LTU threshold.

In the baseline theoretical framework, firms make production and tax reporting decisions given a tax rate on reported profits. Underreporting revenue lowers tax liabilities, but it also implies some resource costs (e.g., keeping two sets of accounting books or foregoing business opportunities). The detection probability increases endogenously with the amount evaded by firms, depending on the interaction between (i) the resources devoted by the tax authority to analyze tax returns and other paper trail generated by business activities ("monitoring effort"); and (ii) the legal reporting requirements and the available technology to process the information generated by business transactions ("enforcement technology"). In line with the quasi-experimental variation created by the LTU, we assume that monitoring effort jumps up discretely at a given level of reported revenue, while the enforcement technology remains constant. This creates an incentive to bunch below the LTU threshold to avoid stricter tax enforcement. We introduce heterogeneity across firms in the resource costs of evasion and in the effectiveness of monitoring effort, which depends on the traceability of firms' transactions. Given that intermediate input sales are easier to trace because they leave more paper trail, the model predicts a stronger bunching response by upstream firms compared to downstream firms, which sell mostly to final consumers. This implies that monitoring effort and the traceability of information trails are complementary.

In the empirical analysis, we use data from financial statements that Spanish firms must submit by law to the Commercial Registry (*Registro Mercantil*). The *Banco de España* compiles and digitalizes this information since 1995, creating a confidential administrative dataset. This dataset contains firm-level information on annual net operating revenue, input expenditures, number of employees, payroll taxes, total value added, and the tax base and liability in the corporate income tax, making it possible to analyze multiple margins of firms' responses to the tax enforcement threshold. The dataset covers more than 80% of registered businesses in Spain with operating revenue between 3 and 9 million euros (the relevant range in our analysis) for the period 1995-2007, during which the LTU threshold remained constant at €6 million.

We document three main empirical findings. First, we find that a significant number of firms report revenues just below the LTU revenue threshold to avoid stricter monitoring. Our bunching estimates indicate that firms reduce their reported revenue by €121,000 on average (about 2% of their total revenue). We estimate the marginal buncher's response to be bounded between €382,000 and €520,000, equivalent to 6.4%-8.7% of total reported revenue. Robustness checks show that the bunching response is not due to other size-dependent policies, and the estimates are robust to different parametric assumptions to estimate the counterfactual distribution.

Second, we study the relationship between monitoring effort and information trails. We define 16 sectors of activity and use the percentage of revenue obtained from sales to final consumers as a proxy for the traceability of transactions made by firms in each sector. We find that the average bunching response is significantly higher in upstream sectors (e.g., heavy manufacturing and wholesale) than in downstream sectors (e.g., retail, restaurants and hotels), suggesting that the effectiveness of additional monitoring effort is higher the easier it is to trace firms' transactions. This finding suggests that information trails and monitoring effort are complements, because it is the interaction of these two elements that deters firms from evading taxes. We show that this complementarity cannot be attributed to a systematic correlation between firm size characteristics, such as number of employees or fixed assets, and the size of the bunching response.

Third, we examine the mechanisms used by firms to avoid stricter tax enforcement. In the presence of multiple taxes (corporate income tax, CIT, value-added tax, VAT, and payroll tax, PRT), firms may evade taxes not only by underreporting revenue but also by overreporting material inputs and underreporting labor costs. To study the nature of the bunching response, we derive predictions on the average relative use of inputs at the threshold depending on whether the bunching response is due to lower output (real response) or underreported revenue (evasion response). Our results show that the average ratio of material expenditures over total revenue shifts down sharply at the LTU threshold. In contrast, the average ratio of the wage bill (net of payroll taxes) as a share of total revenue shifts upward. If bunchers reduced their true output, both ratios would jump up at the threshold, because they are more productive than other firms below the threshold. Hence, while the cross-sectional patterns do not identify causal effects, they are incompatible with a real response. We also rule out the hypothesis that firm size or sectoral composition effects drive these results.

The evidence on reported tax bases suggests that stricter monitoring creates a tax compliance effect on large firms that is equivalent to broadening statutory tax bases. We document a stable one percentage-point gap (5% vs. 6%) in the average reported taxable profit (the tax base of the CIT) and significant jumps in the VAT and payroll tax bases between the two tax-enforcement regimes. These gaps are persistent well beyond the interval around the LTU threshold, suggesting that they cannot be explained only by selection due to the bunching response. Even in the presence of third-party reporting and detailed information requirements, medium-sized and small firms take advantage of their role as fiscal intermediaries appropriate part of the VAT that they should remit to the tax authority, and evade CIT and payroll tax. Overall, the significant tax-reporting regime shift associated to the LTU suggests that this policy is an effective tool to reduce tax evasion.

We extend our theoretical framework to study the net welfare change associated to raising tax monitoring on firms by expanding the tax authority's resources. We show that this policy can be a more efficient way to raise revenue than increasing marginal tax rates if the gains from reducing the resource costs of evasion exceed the administrative cost of the enforcement policy. This framework allows us to examine the mechanisms and conditions under which size-dependent tax enforcement policy, implemented through a LTU, can be welfare improving. Our novel findings indicate that this policy has positive effects on tax compliance, because medium-sized firms evade a substantial amount of taxes even in the presence of extensive information reporting. Applying the theoretical framework to the Spanish case, we provide calculations of the marginal resource costs of evasion indicating the substantial welfare gains of reducing tax evasion of large firms. These calculations also suggest that there would be substantial net welfare gains from extending the scope of the LTU to monitor smaller firms.

This is the first paper to provide credible estimates of the effects of size-dependent tax enforcement policies on firm behavior. Our results are consistent with the insights from the literature on size-dependent policies and firm behavior. One strand of this literature has focused on evasion and avoidance responses at thresholds that determine eligibility to certain taxes, both theoretically (Keen and Mintz, 2004; Dharmapala, Slemrod and Wilson, 2011; Kanbur and Keen, 2014; Bigio and Zilberman, 2011) and empirically (Onji, 2009; Liu and Lockwood, 2015). Other studies have instead analyzed the impact of size-dependent regulations on aggregate productivity, given the pervasive incentives for firms to remain inefficiently small (Guner, Ventura and Xu, 2008; Restuccia and Rogerson, 2008; Garicano, Large and van Reenen, 2016). Our results indicate that firms under low tax monitoring may look smaller in the data than they are in reality because of revenue underreporting, which could have implications for the estimates of the relationship between productivity and firm size.

This study is more broadly related to the literature on tax evasion by firms. Several recent studies have emphasized the importance of third-party information reporting as an effective deterrent against tax evasion (Kopczuk and Slemrod, 2006; Gordon and Li, 2009; Pomeranz, 2015; Kleven, Kreiner and Saez, 2016). Indeed, experimental evidence shows that the individual tax compliance rate is almost perfect when there is third-party reporting, but it is substantially lower for self-reported income (Slemrod, Blumenthal and Christian, 2001; Kleven et al., 2011; IRS, 2012). Nonetheless, it is worth noting that monitoring firms is more complex than monitoring individuals, due to the large amount and variety of transactions that need to be verified jointly to determine the correct tax liabilities in several tax bases.[1] Our finding that monitoring effort and information trails are complementary demonstrates the positive impact of information flows through the production chain on tax compliance, in line with the results obtained by De Paula and Scheinkman (2010) and Pomeranz (2015) in developing economies. However, our results also highlight the difficulty of effectively reducing corporate tax evasion, even in the presence of extensive third-party information reports, without devoting additional monitoring resources to exploit the information flows generated by firms' paper trail.[2]

The empirical techniques used in this paper draw on the rapidly-growing bunching literature, which analyzes agents' responses to thresholds in taxes and regulations.[3] Our

[1] As taxpayers, firms remit corporate income tax and a share of payroll tax. As tax collectors, they withhold income and payroll tax from employees. In most countries, firms also remit value added tax (VAT) payments (Bird and Gendron, 2007; Keen and Lockwood, 2010).

[2] Carrillo, Pomeranz and Singhal (2016) and Slemrod et al. (2015) also document limitations of third-party reporting in two very different contexts, Ecuador and the United States. Both studies find a similar result: firms respond to improved third-party reporting of revenue by evading through less verifiable margins, such as input expenditures, leaving their reporting tax base unchanged.

[3] In the seminal paper of this literature, Saez (2010) exploits kinks—income thresholds at which the marginal tax rate jumps—to identify taxable income elasticities. Kleven (2016) surveys the rapidly-

estimation strategy is closest to the one proposed by Kleven and Waseem (2013), who exploit notches[4]—i.e., income thresholds at which the average tax rate jumps—to identify structural elasticities in the presence of optimization frictions. The novel feature of our setting is that the Spanish LTU generates a notch in enforcement intensity, rather than the tax rate, allowing us to study the effects of tax enforcement policies in isolation.

The rest of the paper is organized as follows. Section 2 presents the theoretical framework. Section 3 describes the empirical strategy and derives the bunching estimators. Section 4 describes the institutional context and the data. Section 5 presents the estimation results. Section 6 discusses welfare and policy implications. Section 7 concludes.

2 Theoretical Framework

We extend the Lucas (1978) model to examine the problem of profit-maximizing firms that can evade taxes and face the risk of being detected and fined. In the baseline setting, firms make production decisions and can underreport their revenue (i.e., evade taxes), which implies incurring a resource cost. The probability of detection depends on the interaction between (i) the *monitoring effort* devoted by the tax authority to examine tax returns and other paper trails generated by firms' transactions and (ii) the existing *enforcement technology* used in this process, including the degree of detail of information-reporting requirements. We use this framework to analyze how firms respond to a discontinuous increase in monitoring effort at an arbitrary revenue threshold. We then extend the model to incorporate heterogeneity across firms in the effectiveness of monitoring effort and in the resource costs of evasion. The model with heterogeneity yields testable predictions on the expected effects of stricter tax enforcement on firms' decisions.

2.1 Corporate Taxation with Risky and Costly Evasion

Consider an economy with a continuum of firms of measure one. Firms produce good y combining tax-deductible inputs x and nondeductible inputs z according to the production function $y = \psi f(x, z)$, where ψ is a productivity parameter and $f(\cdot, \cdot)$ is strictly continuous, increasing and concave in both arguments. Productivity ψ is exogenously distributed over the range $[\underline{\psi}, \overline{\psi}]$ with a smoothly decreasing and convex density $d_0(\psi)$ in the population of firms. Firms purchase deductible and nondeductible inputs in competitive markets at unit cost w and q, respectively, and sell their output at the market price p, which is normalized to unity.

The government levies a proportional tax t on taxable profits $P = y - wx$, so net-of-tax profits with truthful reporting are given by $\Pi = (1-t)P - qz$. Since the tax authority

growing literature using bunching estimation techniques.

[4] Slemrod (2010) provides a general description of notches in tax and regulatory systems, and two recent papers, Best and Kleven (2016) and Kopczuk and Munroe (2015) use bunching estimation techniques to study notches generated by property transaction taxes.

does not perfectly observe all transactions in the economy, firms may attempt to evade taxes by misreporting taxable profits. In the baseline case, firms can underreport their revenue by an amount $u \equiv y - \overline{y} \geq 0$, where $\overline{y}$ is reported revenue[5] and, therefore, reported taxable profits are given by $\overline{P} = (1-t)[\overline{y} - wx]$. The direct and indirect resource costs of evasion are captured with the reduced form $\kappa(u)$, which is an increasing and convex function of concealed revenue.[6] We initially assume this cost function to be the same for all firms, but we relax this assumption later.

The tax authority detects evasion with probability $\delta = \phi h(u)$, where $\phi > 0$ is the monitoring effort parameter, and $h(\cdot)$ is a continuous, increasing and convex function of underreported revenue. The endogenous component, $h(u)$, represents the enforcement technology used to match tax returns among trading partners and to review the paper trail created by information-reporting requirements. Convexity ensures that a larger amount of unreported activity increases the probability of detection more than proportionally, because each reporting inconsistency leaves a new paper trail. The monitoring effort parameter, ϕ, represents the amount of resources devoted by the tax authority to the audit process. It is an exogenous policy choice that encompasses both quantity (e.g., the total number of auditors employed by the tax authority) and quality (e.g., the auditors' skill level and the actual effort exerted during audits).

We model the probability of detection as the interaction between monitoring effort and the available technology to capture the intuition that the two elements are complementary to determine the deterrence effect of tax enforcement policy. Finally, we assume that, when evasion is detected, the tax authority imposes a fine with a penalty rate θ over the amount of tax evaded, on top of the true tax liability.[7]

Firms make production (i.e., demand of inputs x and z) and reporting (i.e., underreported revenue u) decisions in order to maximize expected after-tax profit, given by $\mathbb{E}\Pi = (1-t)[\psi f(x,z) - wx] - qz - \kappa(u) + tu[1 - \phi h(u)(1+\theta)]$. An interior optimum satisfies the following system of first-order conditions:[8]

[5] In subsection 5.3, we discuss the predictions of an extended model in which firms can also evade taxes by misreporting their input costs. We fully derive the extended model in the online appendix.

[6] One example of these resource costs of evasion is the need to maintain two parallel accounting books to keep track of "black" payments made in cash. Tax evading firms may also forego business opportunities by not accepting credit cards or bank payments, given that it is much easier to conceal cash transactions. See Chetty (2009*b*) for a detailed discussion on the economic nature of these resource costs.

[7] For simplicity, we assume that when discrepancies between firms' reported transactions are detected, the authorities uncover the full amount evaded.

[8] The assumption of a convex detection probability is sufficient to ensure the second-order condition for an interior optimum is satisfied.

$$\psi f_x(x,z) = w \tag{1}$$

$$\psi f_z(x,z) = q/(1-t) \tag{2}$$

$$t[1-\phi h(u)(1+\theta)] = \kappa_u(u) + tu(1+\theta)\phi h_u(u) \tag{3}$$

where the term $[1-\phi h(u)(1+\theta)]$ is the expected rate of return of evasion. This system of equations indicates that a positive tax rate has two effects. First, it distorts the choice of inputs, reducing production below the zero-tax optimum. Second, it creates incentives to evade taxes, thereby reducing reported revenue for all firms in equilibrium. Simple comparative statics show that an increase in monitoring effort ϕ leads to a decrease in concealed revenue u without affecting production decisions.

To provide more intuition on firms' incentives to evade taxes, we define the elasticity of detection probability with respect to concealed income as $\varepsilon_{\delta,u} \equiv \phi h_u \cdot u/\delta$, and rewrite the optimal evasion condition (3) as follows[9]

$$1 = \frac{\kappa_u(u)}{t} + (1+\theta)\delta(u)\left[1+\varepsilon_{\delta,u}\right]. \tag{4}$$

The right-hand side of (4) identifies the two mechanisms that contribute to raising tax compliance by firms. The first term shows the disincentive effect created by the presence of marginal resource costs (relative to the marginal benefit of evasion, i.e., the tax rate). The second term represents the deterrence effect generated by the interaction between the tax authority's monitoring effort and the enforcement technology to review the paper trail generated by misreporting behavior.

Given that the production and resource cost functions, $f(\cdot)$ and $\kappa(\cdot)$, are the same for all firms, all the variation in reported revenue $\overline{y}$ is due to differences in productivity ψ. Assuming that monitoring effort ϕ is constant across all firms, there exists a density function of reported revenue $g_0(\overline{y})$ which is smoothly decreasing and convex in its full domain $[\overline{y}_{\min}(\underline{\psi}), \overline{y}_{\max}(\overline{\psi})]$.[10] The theoretical revenue distribution $G_0(\overline{y})$ is smoothly decreasing and convex in firms' productivity, as depicted by the dashed curve in Figure 1a.

Large Taxpayers Unit (LTU): A Tax Enforcement Notch. Assume now that the government provides additional resources to the tax authority in order to create a Large Taxpayers Unit (LTU). This reform raises the monitoring effort without affecting the enforcement technology. Formally, the LTU increases monitoring effort from ϕ_0 to $\phi_1 \equiv \phi_0 + d\phi$ (where $d\phi > 0$) *only* for firms with reported revenue $\overline{y} > y^L$, where y^L denotes the threshold for LTU eligibility. The probability of detection is then given by $\delta =$

[9] This equation is similar to the one derived by Kleven et al. (2011), but obtained from the choice problem of firms, with an additional term to capture the impact of resource costs of evasion.

[10] The specific mapping between the productivity and reported revenue density functions depends on the functional forms of the production function $f(\cdot)$ and the enforcement technology $h(u)$.

$\left[\phi_0 + d\phi \cdot \mathbf{1}\left(\overline{y} > y^L\right)\right] \cdot h(u)$, where $\mathbf{1}(\cdot)$ is the indicator function. The introduction of the LTU creates a tax enforcement notch, meaning that monitoring effort and, consequently, the probability of detection increases discretely at the arbitrary revenue level y^L.

The predicted behavioral response to the tax enforcement notch allows us to classify firms in three groups depending on their exogenous productivity draw: low productivity, bunchers, and high productivity firms. First, "low productivity" firms are those whose incentives do not change with the introduction of the LTU and whose productivity draw is in the range $[\underline{\psi}, \psi^L]$. The upper bound of the set of low productivity firms is defined by the firm with a productivity ψ^L such that its reported revenue is exactly the enforcement threshold y^L in both tax enforcement regimes. The detection probability does not change for those firms and thus their production and reporting decisions are not affected by the introduction of the LTU.

Second, "bunchers" are firms with productivity draws in the range $\left(\psi^L, \psi^M\right]$, which have incentives to avoid being under the LTU scrutiny. The firm with productivity ψ^M, such that its pre-LTU reported revenue is $y^M > y^L$, establishes the upper bound of the bunchers group and is called the marginal buncher. Upon introduction of the LTU, this firm is indifferent between reducing its reported revenue to bunch at y^L (which reduces its expected tax bill but implies resource costs of evasion), or remaining above the threshold facing stricter tax enforcement. Formally, its expected net-of-tax profits are the same in both cases: $\mathbb{E}\Pi_0\left(x, z, u|\phi_0, \psi^M\right) = \mathbb{E}\Pi_1\left(x, z, u|\phi_1, \psi^M\right)$, where subscripts 0 and 1 refer to production and evasion decisions under low and high enforcement intensity, respectively. The set of firms with $\psi \in \left(\psi^L, \psi^M\right)$ consists of those whose expected tax savings associated to bunching are strictly positive, i.e. $\mathbb{E}\Pi_0\left(x, z, u|\phi_0, \psi^M\right) - \mathbb{E}\Pi_1\left(x, z, u|\phi_1, \psi^M\right) > 0$. Hence, the bunchers have incentives to report y^L, generating a mass point point at y^L in the distribution of reported revenue, and a hole with zero mass in the range $\left(y^L, y^M\right)$.

Third, "high productivity" firms are those with productivity draws in the range $\psi \in (\psi^M, \overline{\psi}]$. After the reform, these firms are monitored by the LTU because the costs of bunching (resource costs of evasion plus the increased probability of detection) outweigh the benefits (the expected tax savings under lower enforcement intensity). Under stricter tax enforcement, those firms report higher revenue than they did before the reform. This creates a rightward shift in the distribution of reported revenue above y^L, marginally extending the hole in the post-LTU distribution. The reported revenue distributions pre- and post-LTU are depicted by the solid blue curve in Figure 1a.

To obtain a quantifiable measure of the behavioral response to the tax enforcement notch, we use a first-order approximation to relate the number of bunching firms to the change in the marginal buncher's reported revenue, following the method first proposed by Saez (2010). For analytical simplicity, consider the case in which the LTU raises monitoring effort by a small amount $d\phi = \phi_1 - \phi_0 > 0$, such that the marginal bunching firm adjusts its reported revenue by $d\overline{y}^M$. The adjustment is proportional to $d\psi = \psi^M - \psi^L$, the difference in productivities between the marginal buncher and the firm that locates at the notch before the LTU is introduced. Since there is a direct mapping between the productivity distribution $d_0(\psi)$ and the pre-LTU reported revenue distribution $g_0(\overline{y})$, we can define the number of bunching firms at the threshold as

$$\begin{aligned} B &= \int_{y^L}^{y^L + d\overline{y}^M} g_0(\overline{y})\, d\overline{y} \\ &\approx \frac{1}{2}\left[g_0\left(y^L\right) + g_0\left(y^L + d\overline{y}^M\right)\right] d\overline{y}^M \end{aligned} \tag{5}$$

where $g_0(\overline{y}^L)$ denotes the height of the pre-LTU density distribution at the threshold and we use a trapezoid approximation to account for the fact that the density function $g_0(\overline{y})$ is downward-sloping in the bunching range. The number of bunching firms depends positively on the increase in monitoring effort at the notch and negatively on the resource costs of evasion. Using the approximation in (5), we define the bunching estimator b as the ratio of excess bunching over the average height of the counterfactual density at the LTU threshold, that is,

$$b \equiv \frac{B}{\frac{1}{2}\left[g_0(y^L) + g_0\left(y^L + d\overline{y}^M\right)\right]} \approx d\overline{y}^M. \tag{6}$$

This is equivalent to the standard estimator used in the existing bunching literature, (e.g., Chetty et al., 2011). Geometrically, the estimator approximates the length (in euros) of the interval of reported revenue where bunching firms would have located in the counterfactual scenario without the LTU.

2.2 Heterogeneous Firms

We extend the baseline model to introduce firm heterogeneity along two dimensions: the effectiveness of tax enforcement policy across firms, and the importance of resource costs of evasion. Considering this heterogeneity, the LTU creates different incentives to bunch for firms with the same productivity level, and the model predicts differential increases in reported revenue for high productivity firms (i.e., those monitored by the LTU). As a result, the extended model no longer predicts a hole with zero mass in an interval above the LTU threshold, but only a triangle of missing mass in that interval. In the presence of heterogeneous agents, the bunching method applied to notches allows us to derive estimators of both the average revenue response to the LTU policy attenuated by resource costs of evasion, and the structural response that we would observe in the absence of such costs.

Heterogeneous Effectiveness of Monitoring Effort. We assume now that the effectiveness of tax enforcement policy depends on the traceability of misreported business transactions, which is related to the amount of paper trail. We define paper trail as any document that the tax authority can review in order to detect tax evasion. This includes tax returns, financial statements, credit-card transactions, banking transactions, sales receipts, etc. Among this broad set of documents, the reports on business-to-business transactions play a prominent role in tax enforcement because they can be very easily cross-checked (Kopczuk and Slemrod, 2006; Pomeranz, 2015). For that reason, we assume that an increase in the resources devoted to monitoring activities is more effective at uncovering evasion from firms that sell mainly to other businesses, because their transactions generate substantial paper trail, compared to firms that sell directly to final consumers, where sales generate little or no paper trail.[11]

Under the assumption of heterogeneous effectiveness of tax enforcement, the LTU is more effective at monitoring upstream firms than downstream firms (whose transactions are harder to trace), at each productivity level. As a consequence, upstream firms have stronger incentives to bunch at the LTU threshold, and they also report higher revenue when they belong to the "high productivity" group.

To capture this intuition formally, assume a joint distribution of productivities and monitoring effort with density $\widetilde{g}(\psi, \phi)$ on the domain $(\underline{\psi}, \overline{\psi})$ x $(\phi_0, \overline{\phi})$. The behavioral response for low and high productivity firms at each monitoring effort level is characterized by the set of conditions (1)-(3). Instead, bunchers make an infra-marginal decision taking into account the heterogeneous discrete change of monitoring effort $d\phi$ that they face and their distance from the notch. At each monitoring effort ϕ, firms in the pre-LTU density interval $(y^L, y^L + d\overline{y}_\phi^M)$ have incentives to bunch if they expect tax savings from that action. Note that the revenue response of the marginal buncher, $d\overline{y}_\phi^M$, is increasing in the effective monitoring effort that it faces. We assume that the distribution of monitoring effort effectiveness is such that there is zero mass in the interval $\left(y^L, y^D\right)$, where $y^D = d\overline{y}_{\phi_{min}}^M$ is the pre-LTU revenue level of the marginal buncher facing the smallest effectiveness change. No firm has an incentive to locate in this region, because they could simultaneously reduce reported revenue and increase expected profits.[12] The rest of the bunching interval is given by $\left(y^D, y_{ub}\right)$, where $y_{ub} = d\overline{y}_{\phi_{max}}^M$ is the pre-LTU revenue level

[11] The distinction is particularly stark in tax systems with a value-added tax (VAT) applied using the credit-invoice method, as in the Spanish case. In this context, every business-to-business transaction has to be reported on VAT returns by two different firms, while final consumer sales are only reported by the seller. The amount of paper trail generated by final sales is particularly scarce, or even nonexistent, in the case of cash transactions.

[12] The prediction of a hole in the distribution of firms in a short range just above the LTU threshold only holds when we rule out the possibility that firms face prohibitive resource costs of evasion (e.g., due to honesty or risk aversion), because the costs of misreporting outweigh the benefits in terms of expected tax savings. We relax this assumption in the next subsection.

of the marginal buncher associated to the largest monitoring effort change. Bunching provides expected tax savings only if the change in monitoring effort at the LTU threshold is large enough to overcome differential resource costs $d\kappa(u)$, so only a share of firms in the interval $\left(y^D, y_{ub}\right)$ have incentives to bunch. Since downstream firms experience a smaller increase in effective monitoring effort, they have lower incentives to bunch (for a given productivity level and resource cost function).

With this setup, we can quantify the average reported revenue response from the observed bunching at the LTU threshold. Let $\widetilde{g}_0(\overline{y}, \phi)$ be the joint distribution of reported revenue and monitoring effort when tax enforcement is smooth with constant ϕ, and denote by $g_0(\overline{y}) \equiv \int_\phi \widetilde{g}_0(\overline{y}, \phi) d\phi$ the unconditional reported revenue distribution absent the tax enforcement notch. We can write the excess bunching mass at the threshold as

$$\begin{aligned} B &= \int_\phi \int_{y^L}^{y^L + d\overline{y}_\phi^M} \widetilde{g}_0(\overline{y}, \phi) d\overline{y} d\phi \\ &\approx \frac{1}{2}\left[g_0\left(y^L\right) + g_0\left(y^L + d\overline{y}^M\right)\right] \cdot E[d\overline{y}_\phi^M], \end{aligned} \tag{7}$$

where $E[d\overline{y}_\phi^M]$ is the average response in reported revenue. This term is a weighted average of the marginal buncher's response at each tax enforcement change implied by the LTU. We denote by b_{av} the estimator of the average bunching response in the population, which is the ratio of excess bunching over the counterfactual reported revenue density at the threshold,

$$b_{av} \equiv \frac{B}{\frac{1}{2}\left[g_0\left(y^L\right) + g_0\left(y^L + d\overline{y}^M\right)\right]} \approx E[d\overline{y}_\phi^M]. \tag{8}$$

This estimator measures the average revenue response to the LTU policy attenuated by the costs of evasion. That is, the estimator averages the response of the bunchers and the lack of response of those that do not bunch (*non-bunchers*).

Heterogeneous Resource Costs of Evasion. We now let resource costs of evasion $\kappa(u)$ differ across firms with the same productivity. For instance, the costs of evading a certain amount of taxes might vary depending on the number of employees with whom the employers would need to collude in order to evade (Kleven, Kreiner and Saez, 2016). These costs can also depend on managers' preferences (e.g. risk-aversion and honesty), or on their attention and effort devoted to misreport transactions rather than allocating their skills to productive activities. Besides, evading firms bear the opportunity costs of foregoing business opportunities with trading partners that do not engage in misreporting (Chetty, 2009*a*). Through any of these channels, firms might face different resource cost functions and, therefore, firms with the same productivity have heterogeneous costs of misreporting. In the LTU context, this implies that firms with the same incentives to react face differential costs of adjustment. For some firms with incentives to bunch given their productivity draw ψ and the change in monitoring effort effectiveness $d\phi$, resource costs of evasion can be so large ("prohibitive") that they exceed the benefits of bunching.

Following Kleven and Waseem (2013), we leverage the strong incentives generated by the notch to quantify the structural reported revenue response, i.e. the response that would be observed if evasion were costless. Let $\alpha(\overline{y},\phi)$ denote the proportion of firms that, given the change in monitoring effort that they face, have incentives to bunch but face prohibitively high resource costs of evasion (given their level of reported revenue and monitoring effort effectiveness). For analytical simplicity, we assume that this proportion is constant in the bunching interval associated to the LTU, such that $\alpha(\overline{y},\phi)=\alpha$ for $\overline{y}\in(y^L, y^L+d\overline{y}^M_\phi)$. The presence of prohibitive resource costs is thus the factor that creates mass in the short-range $\left(y^L, y^D\right)$ above the threshold. The excess bunching mass at the LTU threshold is now given by

$$\begin{aligned}B_{adj} &= \int_\phi \int_{y^L}^{y^L+d\overline{y}^M_\phi} [1-\alpha(\overline{y},\phi)]\cdot \widetilde{g}_0(\overline{y},\phi)d\overline{y}d\phi \\ &\approx \frac{1}{2}\left[g_0\left(y^L\right)+g_0\left(y^L+d\overline{y}^M\right)\right]\cdot(1-\alpha)\cdot E[d\overline{y}^M_\phi],\end{aligned} \tag{9}$$

where $E[d\overline{y}^M_\phi]$ is the average reported revenue response at the LTU threshold.

Given (9), the bunching parameter that measures the structural response (i.e., the response that we would observe in the absence of resource costs of evasion) can be written as

$$b_{adj}\equiv\frac{b_{av}}{1-\alpha}\approx E[d\overline{y}^M_\phi]. \tag{10}$$

Expression (10) indicates that the larger the number of bunching firms and the smaller the hole in the bunching region (i.e. the higher the share of non-bunchers) the larger is the structural response to stricter tax enforcement. We discuss in subsection 3.1 how we estimate the lower and upper bounds of this structural response.

3 Empirical Strategy: Bunching Estimators

This section presents the empirical procedure to estimate the reported revenue response of firms to a tax enforcement notch created by the existence of a LTU. To quantify this response, we adapt the techniques from the bunching literature in individual taxation (Saez, 2010; Chetty et al., 2011; Kleven and Waseem, 2013) to estimate the parameters derived in the previous section.

The basic procedure to estimate the reaction of firms to a LTU relies on constructing a counterfactual distribution of reported revenue in the absence of a tax enforcement notch, and comparing it with the observed distribution. To build the counterfactual, we fit a high-degree polynomial to the observed density, excluding an interval around the threshold. We discuss below how the excluded interval is determined. Dividing the data in small bins of width w, we estimate the polynomial regression

$$F_j = \sum_{i=0}^{q} \beta_i \cdot (y_j)^i + \sum_{k=y_{lb}}^{y_{ub}} \gamma_k \cdot \mathbb{1}(y_j = k) + \eta_j, \tag{11}$$

where F_j is the number of firms in bin j, q is the order of the polynomial, y_j is the revenue midpoint of bin j, y_{lb} and y_{ub} are the lower and upper bound of the excluded interval (respectively), and the γ_k's are intercept shifters for each of the bins in the excluded interval. Then, using the estimated coefficients from regression (11), we estimate the counterfactual distribution of reported revenue, that is, $\widehat{F_j} = \sum_{i=0}^{q} \widehat{\beta}_i \cdot (y_j)^i$. This expression excludes the γ_k shifters to ensure that the counterfactual density is smooth around the threshold.

Comparing this counterfactual density to the observed distribution we can estimate the excess bunching mass to the left of the threshold (B), and similarly the missing mass to the right of the threshold (H), given by

$$\widehat{B} = \sum_{j=y_{lb}}^{y^L} \left(F_j - \widehat{F_j}\right) \geq 0 \qquad and \qquad \widehat{H} = \sum_{j=y^L}^{y_{ub}} \left(\widehat{F_j} - F_j\right) \geq 0.$$

We follow the estimation method proposed by Kleven and Waseem (2013) to determine the lower and upper bounds of the excluded region. This procedure imposes that the areas under the counterfactual and observed densities have to be equal, and thus the missing area (H) has to equal the excess mass (B). Implicitly, this is equivalent to assuming that all responses to the tax enforcement notch are on the intensive margin.

The assumption of no extensive-margin responses might be violated. For example, some firms may strategically split up into two or more legally distinct entities to avoid being in the LTU. The resulting firms would not necessarily bunch just under the threshold.[13] In this case, the additional mass to the left of the threshold would push the counterfactual distribution upwards, creating a downward bias in the bunching estimates. However, as some (though not necessarily all) of these firms are coming from the bunching interval $\left(y^L, y_{ub}\right)$, the missing mass would exceed the true bunching mass, introducing an

[13] The LTU's revenue criterion refers to individual legal entities or consolidated fiscal groups (article 121 of the VAT Law), not to corporate groups. Consolidated fiscal groups are quite narrowly defined, requiring that the parent firm owns at least 75% of the subsidiary's capital (art. 67 of Royal Decree 4/2004) and that all firms share the same activity code. Therefore, a firm with annual revenue below €6.01 million would be included in the LTU census if it belongs to a consolidated fiscal group, but not if it is only part of a corporate group. According to the Statistics published by the Spanish Tax Authority from 2004 to 2007, less than 1% of firms in the revenue range €1.5-€6 million are part of the LTU due to belonging to fiscal groups (we do not have firm-level information). Hence, we do not expect this issue to generate a relevant bias in our bunching estimates.

upward bias. There is no reason to believe that these two biases will cancel out exactly, but the size of the net bias relative to the estimates is likely to be small.

Another potential concern is that the method to obtain the counterfactual does not account for the increase in reported revenue by high productivity firms. This intensive-margin response shifts the observed distribution to the right. However, this rightward shift is of second-order importance for the estimation of the counterfactual, because the observed distribution is approximately flat to the right of the bunching segment.[14] Again, it is unlikely that this would create a large bias in the bunching estimator in the current setting.

To obtain the bunching estimates, we first fix the lower bound y_{lb} at the bin where the decreasing trend of the reported revenue density reverts to an increasing rate due to the bunching response. Second, we set the upper bound at $y_{ub} \approx y^L$ and then run regression (11) multiple times, increasing the value of y_{ub} by a small amount after each iteration. When bunching is substantial, the first few iterations yield large estimates of $\widehat{B}$ and small estimates of $\widehat{H}$. This estimation procedure iterates until reaching a value of y_{ub} such that missing and bunching areas converge, i.e $\widehat{B} = \widehat{H}$.[15]

Once we have an estimate for the excess number of bunching firms B and the counterfactual density at the threshold $g_0(\overline{y})$, we can estimate the average bunching parameter b_{av} defined in equation (8). The explicit formula for the estimator is given by

$$\widehat{b}_{av} = \frac{\widehat{B}}{\frac{1}{2}\left(\widehat{F_{y^L}} + \widehat{F_{y_{ub}}}\right)} \tag{12}$$

where $\widehat{F_{y^L}}$ and $\widehat{F_{y_{ub}}}$ are the height of the counterfactual density at the notch and at the upper bound of the bunching interval, respectively. This estimator measures the average reported revenue response of firms in the bunching interval.[16] The magnitude of the response depends on the effectiveness of the policy across firms, attenuated by the costs of their reaction. We estimate the standard errors using bootstrapping.[17]

[14] This point is also noted by Kleven (2016).

[15] In the empirical application there is a finite number of bins, so we impose the condition that the ratio be "close" to one, i.e. $\hat{H}/\hat{B} \in [0.9, 1.1]$. The sensitivity analysis presented section 5 and the appendix shows that all our bunching estimates (b_{av} and b_{adj}) are robust to the choice of the excluded region.

[16] In earlier bunching papers (e.g., Chetty et al., 2011), the b estimator is defined as the ratio of excess bunching mass over the height of the counterfactual density at the threshold, and then the elasticity is calculated separately. In the results presented below, we multiply the estimator of b_{av} times the bin width w in order to obtain a money metric directly.

[17] We thank Michael Best for sharing his Stata code to perform the bootstrapping routine. In all the results shown below, we perform 200 iterations to obtain the standard errors. Using a larger number does not affect any of the results.

3.1 Adjusted Bunching Estimator

In subsection 2.2, we derived the parameter b_{adj} to quantify the structural response that would be observed in the absence of resource costs of evasion. To estimate this structural response empirically, we need to estimate the proportion of firms (α) that, having incentives to bunch given the $d\phi$ that they face at the notch, do not react because the costs of misreporting are higher than the expected tax savings of bunching. According to the model with heterogeneous firms, the only factor preventing a bunching response in the short range (y^L, y^D) is the existence of prohibitively high resource costs of evasion.[18]

For simplicity, we assume that the proportion of non-bunchers, α, is constant in the bunching range.[19] We can then construct the estimator $\widehat{\alpha} \equiv \frac{\int_{y^L}^{y^D} g(\overline{y})d\overline{y}}{\int_{y^L}^{y^D} g_0(\overline{y})d\overline{y}}$. In our empirical application, the upper bound of this interval, y^D, is not uniquely defined because it depends on the change in the effectiveness of monitoring effort, $d\phi$, which is heterogeneous across firms. Hence, to obtain the empirical estimates of α we select several plausible values of y^D and show that the resulting estimates of b_{adj}^{lb} across sectors and years are robust to this choice.[20]

Using the estimate of α, we reweigh the average bunching estimator in (12) to obtain the adjusted bunching estimator

$$\widehat{b}_{adj}^{lb} = \frac{\widehat{b}_{av}}{1 - \widehat{\alpha}}, \tag{13}$$

which provides a lower bound measure of the structural reported revenue response to stricter tax enforcement. This is a lower bound because the proportion of non-bunchers increases with the distance from the notch, rather than being constant as we assume.

The upper bound of the structural response is determined by the convergence point between the counterfactual and the empirical densities, y_{ub}. In the model with heterogeneous monitoring effectiveness, the convergence point identifies the behavioral response of the firm on which the LTU is most effective. Therefore, we interpret the estimator $\widehat{b}_{adj}^{ub} = y_{ub} - y_L$ as an upper bound of the structural response. This is what Kleven and Waseem (2013) call the "convergence method".

[18] In the rest of the bunching interval, (y^D, y_{ub}), both resource costs and a small change in monitoring effort $d\phi$ may prevent bunching behavior, but it is not possible to empirically disentangle the relative importance of the two factors.

[19] This is a conservative assumption, because the importance of resources costs of evasion most likely increases with the distance from the threshold, and so does the proportion of non-bunchers.

[20] These robustness checks are reported in Tables A.7 and A.8 in the online appendix.

4 Institutional Context and Data

4.1 Tax Administration Thresholds: the Spanish LTU

The Spanish tax authority established a Large Taxpayers Unit (LTU) in 1995 to increase monitoring effort on the largest taxpayers, defined as firms with annual operating revenue above €6 million.[21] This eligibility threshold has not been modified since then, and the number of firms in the LTU census (excluding public companies) more than tripled from 11,107 in 1995 to 34,923 in 2007, mainly due to strong economic growth (yearly average of 4%) and inflation (yearly average of 3%) in the period 1995-2007.[22] Over that period, the overall staff of the tax authority–and the LTU in particular–stagnated, while the enforcement technology and IT systems, including the introduction of electronic reporting and modern data management, improved steadily (AEAT, 1995-2007). Thus, the effective enforcement resources per taxpayer available to the LTU could have declined over time, although it is hard hard to quantify the importance of the technological improvements. Firms in the LTU represent only 2% of those that submit a corporate income tax return, but they report about 80% of all taxable profits, 65% of total sales subject to VAT, and they employ 40% of private sector wage earners (AEAT, 1995-2007).

Businesses just above and below the LTU threshold face the same tax schedule on all the major taxes (corporate income, value added and payroll taxes), and the same administrative requirements related to invoicing, accounting and information reporting. The monitoring process on taxpayers works in several stages: first, the tax authority's electronic system processes all detailed the information reported in tax returns to detect inconsistencies or mistakes. Second, the system automatically cross-checks tax returns with other available information such as financial statements or transactions reported by third parties (e.g., trading partners or banks) to detect any discrepancies. Up to this point in the process, the enforcement technology is the same for firms above and below the LTU threshold. Once the system detects an inconsistency on a specific tax return, tax auditors analyze the paper trail available for that firm and request additional clarifications or tax payments using notification letters. If a large inconsistency is detected, the staff may

[21] The threshold was originally set at 1 billion pesetas (€6.010121 million at the official exchange rate). There are two exceptions to the eligibility rule: (i) exporters are always included in the LTU census, and (ii) firms based in two small regions with independent tax authorities (Navarra and País Vasco) are only included if they obtain a large proportion of revenue from sales in the rest of Spain. We provide more details on the LTU eligibility rule, including the role of corporate groups, in online appendix C. In 2006, an additional threshold at €100 million in operating revenue was established to determine eligibility to the Central Office for Large Taxpayers, a select group of the largest firms within the LTU which have access to more individualized legal consultation. We do not observe a bunching response around this threshold.

[22] The impact of economic growth can be seen in the evolution of the total number of firms in the economy, which increased more than twofold in the period of analysis, as shown in Table A.4 in the online appendix.

start a full tax audit.[23]

The key difference we exploit in our empirical strategy is the fact that the LTU has more auditors per taxpayer than the rest of the tax authority, and those auditors have on average higher qualifications and experience. This implies that the LTU makes heavier use of the available information and enforcement technology to detect tax evasion than the rest of the tax authority. This policy setting provides quasi-experimental variation in the monitoring effort (captured by ϕ in the model) faced by large firms, while keeping the information requirements and enforcement technology (captured by $h(u)$) equal for firms around the LTU threshold. Therefore, it allows us to examine firms' responses to a more effective tax enforcement system.

4.2 Data

In the empirical analysis, we use data from financial statements that all Spanish firms must submit by law to the Commercial Registry (*Registro Mercantil*). Firms are also required to reproduce these public official documents on their annual corporate income tax return, so there are no incentives to report different information on the two documents. In fact, auditors from the tax authority routinely cross-check tax returns against published financial statements to detect discrepancies (for example, to ensure that the annual operating revenue figure is the same in both).

The *Banco de España* has compiled and digitalized all the financial statements submitted for the period 1995-2007 to construct a confidential administrative dataset. This micro-dataset has a panel structure and includes the following information for each firm: business name, fiscal identifier, sector of activity (4-digit CNAE-2009 code), Balance Sheet, Profit & Loss Account, and number of employees.[24]

Comparing the number of firms in the *Banco de España* dataset to official statistics of corporate income tax filers for the period 1995-2007, we observe that our dataset contains almost 80% of all firms with reported revenue between €3 and €9 million, the relevant range in the empirical analysis (see Table A.4 in the online appendix for details). The main reason for missing observations is that some firms submit their statements late or on paper form, in which case they may not have been digitalized. The analysis performed by the *Banco de España* statistical division concludes that missing firms are on average less transparent, which may be positively correlated with misreporting behavior. Among

[23]LTU auditors focus exclusively on LTU firms, even though they have access to information on all firms. If they find a discrepancy in the reporting of a non-LTU firm, they can flag it in the system, but they cannot start an audit process against that firm. Therefore, the possibility of tax enforcement spillovers across the LTU threshold is limited.

[24] In an earlier version of this paper (Almunia and Lopez-Rodriguez, 2014) we used a similar dataset from Amadeus, also constructed using financial statements from the Commercial Registry, obtaining almost identical results. We provide further details about the current dataset used, including summary statistics, in online appendix D.

reporting firms, the *Banco de España* dataset identifies some firms as having "unreliable data", often because of rounding issues or inconsistent reporting. We drop these firms (about 7% of the total) and obtain a final dataset of 285,570 firm-year observations with reported revenue between €3 and €9 million from 86,009 unique firms in the period 1995-2007. We expect any bias due to excluding missing and unreliable firms from our sample to go against finding a behavioral response to avoid being in the LTU.[25]

5 Results

We first document and quantify the reported revenue response of Spanish firms to the notch in effective tax enforcement created by the existence of a LTU. Second, we examine the heterogeneity of this response across sectors of activity to test the hypothesis that monitoring effort and the enforcement technology are complements. To do this, we use the relative position of firms in the value chain as a proxy for the effectiveness of monitoring effort in each sector. Finally, we analyze the expected effect of the LTU on tax compliance by analyzing reported tax bases, identifying manipulation in reported input expenditures as an additional channel of tax evasion.

5.1 Bunching Estimates: Reported Revenue Response

Figure 2 shows the distribution of reported revenue for Spanish firms in the period 1995-2007, using administrative micro-data from financial statements collected by *Banco de España*. The graph shows the counterfactual density (dashed line) and the observed density (solid line), overlaid. We focus on firms in the range between €3 and €9 million, such that the LTU threshold is in the center of the graph. There is substantial bunching just below the LTU threshold, indicating that a significant number of firms attempt to avoid stricter tax enforcement by keeping their reported revenue below €6 million.[26]

In order to quantify the excess bunching mass, we estimate the counterfactual distribution by fitting a flexible polynomial, as explained in section 3. For the average bunching estimator, we obtain $\widehat{b}_{av} = 0.121$ (*s.e.* 0.007) which is statistically different from zero at the 1% level. This point estimate implies that firms originally in the bunching interval reduce their reported revenue by about €121,000 on average (approximately 2% of their total reported revenue) in response to the tax enforcement notch. The interpretation of

[25] The Spanish Tax Authority (AEAT) has cooperated with the *Banco de España* to verify that the dataset used in this study is representative of, and bunching results consistent with, the confidential census-level tax return data held by the AEAT. The nature of the statistical exercises done by the AEAT have been reported to this *Journal*, but the strict disclosure policies that apply to the Spanish Tax Authority do not allow the publication of those results outside of its official publications.

[26] There is another modest spike in the distribution at about €4.75 million, associated to a requirement to perform an external audit of the accounts. Since this bunching response is very localized, it does not affect our main estimates. We provide additional discussion in online appendix C, and for the rest of the paper we focus on the response to the LTU threshold.

b_{av} as an average response derives from the assumption that firms are heterogeneous. The estimate of b_{av} is a weighted average of the bunchers' response and the lack of reaction of non-bunchers.

Two features of the distribution of reported revenue deserve additional discussion. First, bunching below the LTU threshold is somewhat diffuse, rather than the single spike predicted by theory. This could be due to the indivisibility of some large transactions, which need to be either reported in full or not reported at all. Another explanation could be the artificial division of firms into multiple entities to avoid crossing the threshold. As discussed in section 3, the sign of the bias created by this extensive-margin response is theoretically ambiguous. In the sensitivity tests reported below, bunching estimates are robust to a wide range of parametric choices, so this potential bias is a minor concern.

Second, we do not observe a "hole" with zero mass to the right of the LTU threshold. Instead, there is a substantial number of firms reporting revenue just above y^L. The model with heterogeneous firms predicts the existence of mass in the bunching interval for two reasons: heterogeneity in the change in monitoring effectiveness ($d\phi$), and variation in the resource costs function $\kappa(u)$ across firms. Taken together, these two features lead to the existence of a triangle of missing mass to the right of the LTU threshold, which can be seen empirically comparing the observed and counterfactual distributions of reported revenue in Figure 2.

The empirical procedure presented in subsection 3.1 allows us to estimate bounds for the structural response to the policy. Using the adjusted bunching estimator, we obtain $\widehat{b}^{lb}_{adj} = 0.382$ (s.e. 0.036). Under the assumption that resource costs of evasion are heterogeneous across firms and increasing with the distance to the threshold, this represents a lower bound of the structural response. The upper bound is given by the point of convergence between the observed and counterfactual densities. Convergence occurs at €6.53 million, so the upper bound estimate is $\widehat{b}^{ub}_{adj} = 0.52$. These point estimates imply that the structural reported revenue response to avoid being monitored by the LTU is bounded between €382,000 and €520,000 (that is, 6.4% to 8.7% of total reported revenue for firms around the LTU notch). The gap between the upper and lower bounds of the structural response suggests that there is more uncertainty in the estimation of this parameter than in the case of the average bunching estimator (b_{av}).

Robustness. We examine the robustness of the bunching estimates reported above in several ways. First, the observed response could be affected by other size-dependent policies, such as a corporate income tax benefit that offers a 5 percentage-point lower tax rate to small firms.[27] We do not find any evidence of bunching in response to this tax

[27] The eligibility threshold for this tax benefit changed over time, from €1.5 million in 1995 to €8 million in 2007, as shown in Table A.2 in the online appendix. The distribution of reported revenue

break, even when we only include firms with positive taxable profits in the estimation. The lack of reaction to such a large reduction in the corporate income tax rate is remarkable in a context where firms respond strongly to a discontinuity in tax enforcement intensity. This evidence indicates that a relevant share of Spanish firms perceives that being under the LTU has a potentially large cost.[28]

Second, pooling several annual cross-sections together increases the effective sample size allowing us to obtain very precise estimates, but it could mask differences in the response across years. Table 1 shows that bunching estimates are statistically significant in every year for the period 1995-2007.[29] There is some variation in bunching across years, partly due to the smaller sample size in the annual subsamples. Bunching is somewhat stronger in the earlier years, which could be due to the fact that the LTU threshold remained fixed in nominal terms throughout the entire period and hence the effective change in monitoring effort $d\phi$ may have declined over time, as suggested in section 4.[30] In order to obtain more precise estimates, for the remainder of the paper we focus on the pooled 1995-2007 dataset.

Third, in the results reported above, we set the lower bound of the excluded region equal to $y_{lb} = 5.7$ and the order of the polynomial to construct the counterfactual equal to 5. We estimate the proportion (α) of non-bunching firms using the interval $\left(y^L, y^D\right) = (6.01, 6.15)$ as the baseline case. An exhaustive sensitivity analysis shows that the results are robust to a wide range of values for the lower bound (y_{lb}), the order of the polynomial (q), the bin width (w) and the upper bound of the non-bunchers' interval (y^D).[31] In the empirical exercises in the next subsection, we focus on the average bunching estimator ($\widehat{b}_{av}$) because the attenuated response is the most relevant for policy and welfare discussions, given that resource costs of evasion are not expected to decline over time.

under each of the thresholds is shown in Figure A.1.

[28] Devereux, Liu and Loretz (2014) study firms' responses to corporate tax kinks in the UK and find significant bunching. In their setting, kinks are set in terms of taxable profits, whereas in the Spanish case the tax incentive depends on operating revenue, which is considered less easily manipulable (Best et al., 2015; Carrillo, Pomeranz and Singhal, 2016). This makes the strong bunching in reported revenue that we observe at the LTU threshold even more remarkable.

[29] The annual histograms are shown in Figure A.2 in the online appendix. In an earlier working paper version (Almunia and Lopez-Rodriguez, 2014), we reject the possibility that these patterns emerge because of repeated bunching, i.e. a small proportion of firms stay in the bunching interval for many consecutive years. The conclusion from the analysis is that the static bunching estimates are not systematically biased due to persistent bunching behavior, because there is only evidence of short-term persistence for a proportion of bunching firms (i.e. some firms remain at most 2-3 years in the bunching interval).

[30] The number of bunching firms significantly increased over time because of the strong nominal growth of the economy (see section 4), but the relative impact of the LTU on the firm size distribution, i.e. the bunching estimates, declines smoothly over time as shown in Table 1.

[31] The results of these robustness checks are reported in online appendix Tables A.6, A.7 and A.8.

5.2 Heterogeneous Responses: Complementarity Result

In the model with heterogeneous firms, we assume that the effectiveness of monitoring effort (ϕ) depends on the traceability of a firm's transactions. According to the theoretical predictions, we expect a larger response to the LTU threshold for upstream firms (which sell mostly to other firms) than for downstream firms (which sell mostly to final consumers). In practice, it is relatively easy to cross-check tax returns to detect misreported intermediate input sales because the buying firm has an incentive to record its expenses to claim tax credits. In the case of final sales, the consumer has no incentive to keep a receipt and therefore it is significantly harder to cross-check those transactions against other information sources, especially when they are made in cash.[32]

To study heterogeneous responses to the LTU threshold empirically, we define 16 sectors of activity.[33] Since we do not have transaction-level data for each firm, we use the percentage of sales made to final consumers in each sector (from input-output tables[34]) as a proxy for the traceability of sales made by firms in that sector. Figure 3 plots the percentage of final consumer sales (in the horizontal axis) against the average bunching estimates by sector (vertical axis). The relationship is downward-sloping, suggesting that the incentive to remain under the LTU threshold is stronger in sectors where a low percentage of sales is made to final consumers. On the top-left corner we observe sectors such as specialized construction activities, transportation, and metal and equipment manufacturing. Firms in these sectors sell less than 10% of their output to final consumers, and they all present high average bunching estimates, between 0.12 and 0.20. On the bottom-right corner, we find retailers and restaurants and hotels. These businesses obtain more than 80% of their revenue from sales to final consumers and they feature a much lower bunching response, between 0.07 and 0.09.[35]

The negative correlation between a high share of final consumer sales and the size of the bunching response at the enforcement notch is consistent with the predictions of our theoretical framework. The empirical result suggests that the deterrence effect associated to higher monitoring resources is most effective for firms whose misreported transactions are easier to detect. In contrast, the increase in monitoring resources is less binding for firms that sell mostly to final consumers. This does not inform us about the *level* of tax

[32] Naritomi (2015) studies an innovative policy implemented in the state of Sao Paulo (Brazil), where consumers were incentivized to keep receipts from all their purchases (by entering a lottery and receiving a partial income tax rebate). Consistently with our framework, she finds that the program had a much stronger effect on the sales reported by retail sector firms compared to wholesalers.

[33] Details about how we define each of the sectors can be found in online appendix D.

[34] We use the input-output tables for the Spanish economy in year 2000 published by the Institute of National Statistics (INE).

[35] The counterfactual and empirical distributions of revenue in all sectors are shown in the online appendix Figure A.3, and all the point estimates are reported in Table 2.

evasion in different sectors, as it could well be that the actual amount evaded is higher for restaurants and hotels than for metal manufacturers. The key insight is that, in sectors with a high share of final sales, even a highly-skilled team of LTU auditors may be unable to detect evasion using standard methods such as information cross-checks. In conclusion, the empirical evidence suggests that monitoring effort and the existence of a paper trail are complements. The deterrence component that contributes to higher tax compliance derives from the interaction of these two elements, rather than from each of them independently. Finally, note that complementarity is a local result: if the level of evasion is very low (or zero), additional monitoring effort will be ineffective.

Robustness. One potential concern is that other firm size characteristics, such as the number of employees or the level of fixed assets, might prevent firms from engaging in tax evasion. Kleven, Kreiner and Saez (2016) make this point theoretically, arguing that larger and more complex firms are less likely to reach a colluding agreement to evade taxes, as there is a higher chance that one of the employees may act as a "whistleblower". Another potential concern is that the resource cost of evasion $\kappa(u)$ might vary systematically across sectors in a way that could be correlated both with bunching and the share of final sales.[36]

Table 3a reports a set of OLS regressions of average bunching estimates by sector on the share of final sales. In the simple regression (unconditional) case, the coefficient is -0.072 (s.e. 0.028), which implies that a 10 percentage-point increase in the share of final sales is associated with an average reported revenue response to the LTU threshold €7,200 lower in that sector. Controlling for the median number of employees and the median tangible fixed assets (a proxy for complexity), has no effect on the size nor the statistical significance of the coefficient of interest. When we also control for the share of non-bunchers (α), the coefficient declines slightly in absolute value to -0.054 (s.e. 0.023), but it remains statistically and economically significant. Table 3b reports the results of weighted least squares (WLS) on the same specifications, using the inverse of the variance of the bunching estimates as weights. The coefficient on the share of final sales decreases in absolute value to -0.049, but it remains strongly significant and robust to using different sets of controls. In additional robustness checks reported in the online appendix, we show that the raw correlation between median employees (or median fixed assets) with average bunching is close to zero and statistically insignificant (Figure A.4 in the online appendix). Moreover, Figure A.5 and Table A.9 show that the complementarity result holds separately for the subsamples of firms below and above the median of the two firm size characteristics. Finally, the result is robust to different choices of the bin width in the bunching estimation (Figure A.6).

These tests suggest that the complementarity result is not due to a systematic relation

[36] We thank an anonymous referee for raising this point.

between firm size characteristics and bunching. The share of final sales is a robust determinant of the size of the bunching response in each sector, supporting the theoretical prediction that monitoring effort and the enforcement technology are complements.

5.3 LTU Effectiveness and the Effect on Tax Compliance

In this subsection, we examine the nature of the behavioral response around the LTU threshold and the impact of the policy on the behavior of firms monitored by the LTU. This is a first step towards analyzing the welfare implications of the policy, but it is challenging from an empirical point of view because bunching is an endogenous response, so we cannot apply a standard regression discontinuity design.

To examine the effect of the LTU on tax compliance, we extend our theoretical framework to allow firms to evade taxes by misreporting their input expenditures (besides misreporting their revenue) in the presence of multiple taxes, namely corporate income tax (CIT), value-added tax (VAT) and payroll tax (PRT). Enabling firms to misreport their inputs yields testable predictions on whether the bunching response is on average due to a change in real production or simply to tax evasion. We summarize the predictions of this extended model below (the full derivation is in the online appendix) and test them empirically using information on reported input expenditures and tax bases. We use these tests to rule out mechanisms that are inconsistent with theory, rather than to identify causal effects.

Theoretical Predictions with Multiple Taxes. The extended model with multiple taxes shows that firms have incentives to underreport revenue and overreport material expenditures in order to evade both VAT and CIT. Similarly, there is an incentive to underreport labor expenditures to evade PRT and avoid regulatory costs of labor.[37] Since the probability of detection is higher under the LTU, firms above the threshold have a smaller incentive to evade taxes and they report larger amounts in each of the three tax bases. To study the mechanisms associated to the bunching response, we consider as outcome variables the average reported ratios of tax-deductible input expenditures (materials and labor) as a fraction of operating revenue. We evaluate the predictions of the model under three scenarios:[38]

(a) *Real production response*: bunching firms that reduce production have a higher

[37] Underreporting labor expenditures increases corporate tax liabilities, but this can be compensated by the tax savings on the payroll tax. During the period under study, the statutory payroll tax in Spain was 38% (including both the employer's and the employee's shares), compared to a corporate income tax rate that declined from 35% to 30%. Moreover, keeping reported salaries low and paying part under the table protects firms against future negative shocks, because there is downward nominal wage rigidity. There is theoretical support for wage underreporting via collusion agreements between employers and employees in Yaniv (1988), and pervasive empirical evidence of this practice in developing countries (Kumler, Verhoogen and Frias, 2015; Best, 2013) and even in the US (Slemrod and Gillitzer, 2014).

[38] See online appendix B for the mathematical derivation of these predictions.

productivity draw (ψ) than those originally located below the threshold, so they need fewer inputs to produce the same amount of output. As a result, average material and labor expenditures as a fraction of revenue follow a downward slope in the interval (y_{lb}, y^L) due to the increasing share of bunchers, and both ratios shift upward at the threshold.

(b) *Evasion via revenue underreporting*: bunchers demand the same inputs (and produce the same output) as they would without an LTU, but they do not report a fraction of their revenue. This response mechanically increases both reported input ratios, yielding an upward trend in both ratios in the interval (y_{lb}, y^L) and a downward shift at the threshold.

(c) *Evasion via revenue and input misreporting*: firms with a larger scope to misreport inputs to minimize their tax liability (i.e., overreporting materials and underreporting labor costs) have stronger incentives to bunch to avoid being monitored by the LTU, because they obtain larger tax savings in expectation. Combined with revenue underreporting, this leads to an upward trend in the materials ratio in the interval (y_{lb}, y^L), and a downward shift at the threshold. The prediction is exactly the opposite for the ratio of labor expenses over reported revenue.

Empirical Evidence: Reported Input Expenditures. Figure 4a plots the average reported ratio of material expenditures to reported revenue (vertical axis) against reported revenue (horizontal axis) for the period 1995-2007.[39] The ratio follows an upward slope in the reported revenue range between €3 and €6 million with a concave shape, indicating that firms with higher revenue use an increasingly larger proportion of material inputs. The relative use of material inputs shifts sharply downwards at the LTU threshold by about two percentage points (from 79% to 77%). Figure 4b shows a similar plot for the reported ratio of the total wage bill (net of payroll taxes) over revenue. The pattern in this case is reversed: the ratio slopes down smoothly with an upward jump of about one percentage point (from 10.5% to 11.5%) at the threshold.

Overall, this evidence is consistent with the scenario of evasion response with both revenue and input misreporting. Under this hypothesis, bunching firms conceal part of their revenue to avoid the stricter tax enforcement of the LTU, which gives them more leeway to misreport their input expenditures. One potential alternative explanation for the observed patterns could be that labor-intensive (or capital-intensive) firms are less likely to bunch due to prohibitive resource costs of evasion. Indeed, if such costs were systematically correlated with the number of employees or the level of fixed assets, the

[39] To avoid the spurious effect of extreme values, we winsorize observations in the top and bottom 1% of the outcome variable. That is, we set extreme values in each €1-million interval equal to the 1st and 99th percentile. In addition, we use wider bins than in the histograms shown before (€120,202 instead of €60,101) to reduce the amount of noise in the bin averages, although the results are robust to different bin widths as shown in Figure A.13 in the online appendix. We do not adjust for inflation because the outcome variables are ratios of two nominal amounts.

bunching response would create discontinuities in the input ratios due to composition effects. We discard this hypothesis by showing that the bunching response is very similar in each of the four quartiles of the employment distribution (or the tangible fixed assets distribution), as shown in Figures A.7 and A.8 in the online appendix. Moreover, Figures A.9-A.12 show that the breaks in average reported material and labor expenditures at the LTU threshold are not driven by a particular subset of labor or capital-intensive firms. Instead, we observe clearly differentiated patterns in these variables on each side the LTU threshold for all quartiles of both variables.[40]

Empirical Evidence: Reported Tax Bases. We study the implications for tax compliance by comparing reported tax bases by firms monitored by the LTU and those below the eligibility threshold. To do this, we exploit the richness of our dataset, which includes the reported tax bases and tax liabilities of the CIT and PRT, and a proxy for the tax base of the VAT.

Figure 5a shows average taxable profits (the tax base of the CIT) as a percentage of reported revenue (left) and in million euros (right).[41] In the left panel, we observe an upward shift in the average reported taxable profit margin from 5% below the threshold to 6% above. The average taxable profit margin is remarkably stable on both sides of the threshold, even far away from the interval affected by selection due to bunching (around €5.5-€6.5 million). This suggests that there is a "reporting regime shift" at the point where firms become eligible for the LTU, which cannot be driven only by selection. This 1pp gap in taxable profit margin is broadly consistent with the overreporting of materials and the underreporting of labor expenditures documented above. In the right panel, we see the corresponding upward shift in the average reported taxable profit, which jumps by about €60,000 at the LTU notch. This evidence is consistent with the reporting effect associated to the LTU, as there is no change in the tax schedule at the threshold.

Figure 5b shows average reported value added as a percentage of revenue (left) and in million euros (right). This variable is a good proxy for the tax base of the VAT at the firm level.[42] We observe a large break at the LTU threshold, with an upward shift from approximately 20% to 23%, indicating that firms monitored by the LTU report a substantially higher value-added tax base. As in the previous graph, the patterns are very different even far from the LTU threshold, suggesting that they are not due to selection

[40] Additionally, Figure A.14 plots average reported inputs for two polar cases: wholesale (low labor intensity, high bunching) and retail (high labor intensity, low bunching). Within both sectors, we observe large breaks in the distribution of reported inputs, in large with the general patterns.

[41] We can only calculate the taxable profit margin for firms with positive taxable profits, because firms with negative profits report a zero tax liability. This implies restricting the sample to 80% of its size. All results presented in the paper are robust to using only the subsample of firms with positive tax liability.

[42] This measure of value added is calculated by the *Banco de España* as the difference between operating revenue and expenses that are considered part of the VAT base.

created by the bunching response. By failing to report the correct tax base, firms outside the scope of the LTU are taking advantage of their role as fiscal intermediaries to keep part of the tax revenue that they should remit to the tax authority.

The tax base of the payroll tax is best approximated by the net wage bill. In Spain, the payroll tax consists of a combination of employer contributions (31% of the net wage in the period we study) and employee contributions (7% of the net wage). In the data, we separately observe the net wage bill and the employers' part of the payroll tax. As discussed above, Figure 4b shows a shift in average net wages as a fraction of reported revenue from 10.5% to 11.5% at the LTU threshold. Once again, there are two significantly different patterns on either side of the notch, suggesting that firms monitored by the LTU have a lower ability to misreport their payroll tax base than those below.[43]

Taking the evidence on the three tax bases together, the empirical patterns depict a persistent tax regime shift associated to an arbitrary nominal threshold over a long period of time (1995-2007). This evidence suggests that the LTU effectively increases tax compliance on the subset of firms that it monitors. Firms below the LTU threshold, including both bunchers and low productivity firms, seem to have more leeway to evade taxes. On average, firms under stricter tax enforcement engage in less misreporting and the LTU policy is therefore equivalent to a broadening of reported tax bases. At a general level, we conclude that the deterrence channel is more effective at reducing evasion when the tax authority has more resources to monitor the information trails making full use of the enforcement technology.

6 Welfare Analysis and Policy Implications

We use the baseline model to examine analytically the welfare effects of increasing tax enforcement on firms. First, we analyze the net welfare change from increasing monitoring effort across firms and returning the additional revenue lump sum to all taxpayers. Second, we examine the mechanisms and conditions under which size-dependent tax enforcement is a welfare-improving policy. Finally, we use the analytical insights and the empirical results on the Spanish LTU to illustrate the potential welfare gains of extending stricter tax enforcement to medium-sized and small firms.

6.1 Welfare Effects of Increasing Tax Enforcement

Consider the baseline economy introduced in section 2. Without loss of generality, assume that each firm is owned by one individual whose total income is the net-of-tax profit of the firm. The government devotes fixed costs F to create a tax authority in charge of

[43] As a robustness check, Figures A.15 and A.16 in the online appendix replicate the evidence for reported CIT and VAT bases for five broad sectors of activity. These results indicate that the tax-reporting regime shift observed in the aggregate data is not due to a systematic correlation between the strength of the bunching response and the sectoral composition of that response.

tax enforcement with a given technology. The tax authority spends $c(\phi)$ resources per taxpayer to employ qualified staff to examine the discrepancies detected by the enforcement technology and undertake tax audits to uncover evasion. The administrative costs of tax enforcement, $AC(\phi) = F + c(\phi)$, are increasing and convex ($c_\phi, c_{\phi\phi} > 0$) in monitoring effort, which is assumed constant across firms in this baseline model. The social welfare function that aggregates expected profits by firms and expected tax revenue by the government, net of administrative costs of enforcement, is given by

$$\begin{aligned}\mathbb{E}W \equiv & \int_{\overline{y}_{\min}}^{\overline{y}_{\max}} \{(1-t)P - qz - \kappa(u) + tur\} \cdot g_0(\overline{y}) d\overline{y} \\ & + \int_{\overline{y}_{\min}}^{\overline{y}_{\max}} \{t(P-u) + \phi h(u) tu(1+\theta)\} \cdot g_0(\overline{y}) d\overline{y} - AC(\phi). \end{aligned} \tag{14}$$

By the envelope condition, the increase in monitoring effort has only a mechanical first-order effect on the firms' expected profits, because firms have already made their real and reporting choices. The marginal change in welfare in response to an increase in monitoring effort is, therefore,

$$\begin{aligned}\frac{d\mathbb{E}W}{d\phi} &= \int_{\overline{y}_{\min}}^{\overline{y}_{\max}} \left[t\frac{d\overline{P}}{d\phi} + t\phi(1+\theta)\left(\frac{\partial h}{\partial u}\frac{du}{d\phi}u + h(u)\frac{du}{d\phi}\right)\right] \cdot g_0(\overline{y}) d\overline{y} - c_\phi(\phi) \\ &= \int_{\overline{y}_{\min}}^{\overline{y}_{\max}} \left[t\frac{dP}{d\phi} - \kappa_u(u)\frac{du}{d\phi}\right] \cdot g_0(\overline{y}) d\overline{y} - c_\phi(\phi), \end{aligned} \tag{15}$$

where the second equality follows from firms' first order condition of net-of-tax profit maximization with respect to concealed revenue. Comparative statics indicate that the increase in monitoring effort without changing the relative cost of inputs does not modify the demand for inputs and thus it does not affect real production, $dP/d\phi = 0$.[44] Arranging terms, we obtain a condition showing that increasing tax enforcement is a welfare-improving policy when the marginal welfare gains from a lower level of tax evasion are larger than the marginal administrative costs of achieving that level, such that

$$\int_{\overline{y}_{min}}^{\overline{y}_{max}} \left[-\kappa_u(u) \cdot \frac{du}{d\phi}\right] \cdot g_0(\overline{y})\, d\overline{y} > c_\phi(\phi). \tag{16}$$

This expression indicates that the welfare gains of increasing tax enforcement depend positively on i) the marginal resource costs of evasion, $\kappa_u(u)$, because the more costly evasion is the larger the gains of reducing it; and ii) the reduction in misreported income due to the increase in monitoring effort. The welfare gains are negatively related with the initial monitoring effort, ϕ, due to convex administrative costs, indicating that there

[44] Our empirical results for the Spanish LTU support this prediction as we do not find any evidence of real responses (i.e. fall of production) to the change in monitoring effort.

are decreasing returns to increasing tax enforcement.

Condition (16) is derived according to the procedure followed in Chetty (2009*a*) to obtain the generalization of Feldstein (1999)'s condition to examine the excess burden of taxation. In contrast to raising marginal tax rates, increasing monitoring effort does not affect relative input prices and thus it broadens reported tax bases without distorting production. Indeed, a higher monitoring effort lowers the returns to misreporting tax bases, yielding direct welfare gains through the reduction in resource costs and a redistributive effect by transferring income among taxpayers. Note that the welfare gains from raising tax enforcement could be even larger if the extra revenues obtained were used to fund valuable public spending or to decrease corporate income tax rates. The latter possibility would reduce the excess burden of taxation, justifying an expansion of the tax agency's resources as discussed in Slemrod and Yitzhaki (1987).

The conditions derived in the extended model with heterogeneous firms (discussed in section 2.2) show that firms respond differently to increases in tax enforcement because i) monitoring effort is more effective on firms whose transactions are easier to trace; and ii) the costs of misreporting transactions can be heterogeneous across firms, to the point that some taxpayers do not react to avoid higher enforcement. Given these insights and condition (16), we can infer that welfare gains from increasing tax enforcement would be higher in economies with a large proportion of upstream firms because their misreported transactions are more likely to be detected.

6.2 Size-Dependent Tax Enforcement

Most governments around the world have established LTUs within their tax agencies to strengthen tax administration and improve tax compliance by the largest taxpayers, often by recommendation of the IMF (IMF, 2002). Despite the widespread adoption of this policy in recent decades, there have been very few attempts to evaluate its welfare effects beyond noting that it has been generally successful at increasing total tax revenue. In practice, there is variation across countries in the specific eligibility criteria, which usually include either total revenue, number of employees, or both. However, the overarching feature of LTUs is that they lead to size-dependent tax enforcement intensity by applying higher monitoring effort per taxpayer to a subset of large firms (OECD, 2011).

We develop a framework to analyze the welfare implications of the policy, taking into account the local distortions generated by bunching to avoid the LTU and also the behavioral responses within the LTU. Using this framework and the estimates from our empirical application, we do a simple welfare calculation of the effects of the Spanish LTU to illustrate the potential welfare gains associated to extending stricter tax enforcement to smaller firms.

Given the social welfare function defined by (14), the welfare change induced by the LTU with an eligibility threshold y^L is given by

$$\begin{aligned}\Delta\mathbb{E}W &= \int_{y^L}^{y^L+d\overline{y}^M}\left[\Delta\Pi^R + \Delta u \cdot \ \kappa_u\left(u\right)\right]\cdot(1-\alpha(\overline{y}))\cdot g_0(\overline{y})d\overline{y} + \\ &+ \int_{y^L}^{y^L+d\overline{y}^M}\left[\Delta u \cdot \ \kappa_u\left(u\right)\right]\cdot\alpha(\overline{y})\cdot g_0(\overline{y})d\overline{y} + \\ &+ \int_{y^L+d\overline{y}^M}^{\overline{y}^{\max}}\left[\Delta u \cdot \ \kappa_u\left(u\right)\right]\cdot g_0(\overline{y})d\overline{y} + \\ &+ \int_{y^L+d\overline{y}^M}^{\overline{y}^{\max}}\left[c(\phi_1)-c(\phi_0)\right]\cdot g_0(\overline{y})d\overline{y}\end{aligned} \tag{17}$$

where $\Delta\Pi^R$ is the change in gross profits (i.e., $\Pi^R = P - qz$) and Δu is the change in reported tax base, both changes induced by the increase in monitoring effort; and $\alpha(\overline{y})$ is the proportion of non-optimizers at each level of pre-LTU reported revenue.

The welfare change associated to a size-dependent tax enforcement policy is thus the result of four components. The first one captures the local welfare loss created by the bunchers' reaction. Bunchers reduce their reported revenue creating welfare losses through the additional resource costs (evasion channel) and/or the reduction in output (real channel). The second component measures the welfare gains due to the adjustment costs that prevent some firms from bunching. These non-nonbunchers report larger tax bases, reducing their resource costs of evasion. Similarly, the third component captures the positive welfare gains of the LTU's deterrence on high productivity firms, which report larger tax bases and hence incur lower resource costs. Although firms monitored by the LTU are worse off because they pay more taxes, this income is a transfer to the government that does not affect aggregate welfare. The key mechanism to obtain welfare gains is the reduction in the resource costs associated to evasion. The final component is the additional administrative costs needed to increase monitoring effort for firms within the LTU. Overall, the policy is welfare improving when the gains from the lower resource costs of evasion born by firms monitored by the LTU, net of the administrative costs, are higher than the local welfare costs created by the bunchers' reaction.

6.3 Welfare Calculations: Application to the Spanish LTU

We evaluate now the welfare effect of expanding the scope of the LTU to one additional firm, using the insights from equations (16) and (17) and the empirical results from our application for the Spanish LTU. Framing the question as a marginal reform from the status quo is a simple way to explore the existence of potential marginal net welfare gains of the policy.[45] Given our analysis, extending the LTU's monitoring effort to an additional firm would be welfare improving if the reduction in resource costs is higher than the additional administrative costs required, given some local welfare losses associated to bunching. The existence of substantial net welfare gains at the margin would indicate the efficiency of spending additional monitoring resources on smaller firms, and also suggest the presence of sizable welfare gains in the interior of the LTU.

The marginal resource cost of evasion is a critical measure to evaluate the welfare gains of extending the policy but, as discussed by Chetty (2009*a*), its estimation is quite challenging. We propose a method to put bounds on this measure by looking at tax reporting behavior by firms in the short-range region above the LTU notch $\left(y^L, y^D\right)$, defined in section 2.2, where firms only locate if they face large adjustment costs. Assuming that these costs are pure resource costs of evasion,[46] we estimate that non-bunchers in the region $\left(y^L, y^D\right)$ report on average an additional one percentage point in taxable profit margin than firms under lower monitoring effort.[47] Over the period of analysis, firms in that region report on average an additional €60,000 of taxable profits which, at a 32.5% tax rate, result in €19,500 of taxes that they could avoid paying by simply reporting €1 less in total revenue. This suggests that these firms bear resource costs of at least 5.9% of their gross income (note that average gross taxable profit for firms around the LTU threshold is €330,000), which prevent them from reacting. Given the shift in the average taxable profits, the inclusion of an additional firm in the LTU would yield a welfare gain of €3,540, due to the reduction in resource costs of evasion.

As an alternative, taking also into account the potential tax savings in multiple taxes such as the PRT (1pp of tax base shift under a 38% tax rate representing €22,800 of tax liability) and the VAT (3pp of tax base shift at 11.5% of effective tax rate over the period resulting in an additional €20,700 tax remittance), the marginal resource costs could be as large as 19% of their gross income. Given the average increase in reported tax bases of approximately €300,000, the welfare gain generated by marginally extending the LTU

[45] Conducting a welfare evaluation of the LTU policy as a whole would require making very strong assumptions about how the local estimates obtained at the threshold extrapolate to the full range of firms. That exercise would incorporate a much larger degree of uncertainty than the one we develop here. For that reason, we just provide back-of-the-envelope calculations of the marginal benefit of extending the LTU and the inferred revenue threshold where these benefits are exhausted. See Kanbur and Keen (2014) and Bigio and Zilberman (2011) for an analytical discussion on the complexity of determining optimal tax thresholds.

[46] If firms do not react because of misperception about tax enforcement strategies, the proposed bounds would be biased upwards. However, the proposed measure captures other costs such as managerial effort to understand the effects of the policy, resources devoted to keeping double accounting, violation of ethical principles, or foregoing business opportunities to maintain an inefficiently small size.

[47] As discussed in section 5.3, the parallel reporting regime shift of 1pp in the corporate income tax base is documented well beyond the revenue region affected by selection. Besides, evidence suggests that on average bunchers mimic tax reporting by small firms, making the 1pp gap stable in the whole revenue interval we study.

would be up to €57,000 in the most optimistic scenario. The bounds on the marginal resource costs of evasion (5.9%-19%) put our estimates in the same order of magnitude as the ones obtained by Gorodnichenko, Martinez-Vazquez and Peter (2009) in a study about the Russian flat tax reform.

To quantify the marginal administrative cost of including one additional firm in the LTU, we need to estimate the average additional cost per taxpayer associated to the monitoring effort of the LTU, that is $c(\phi_1) - c(\phi_0)$. According to official data published by the Spanish tax agency (AEAT, 1995-2007; Tribunal de Cuentas, 2008), over the period of analysis the LTU has been endowed with 125 highly-qualified tax auditors earning an average annual compensation between €60,000 and €80,000.[48] This group of auditors is in charge of monitoring the approximately 30,000 firms in the LTU Census. Hence, we can approximate the additional cost per taxpayer in the LTU to be in the range of €250-€333.[49] The small magnitude of the marginal cost per taxpayer illustrates the potential gains of exploiting the fixed-cost structure of the tax agency by extending strict monitoring effort to smaller firms below the current LTU eligibility threshold.[50]

Our simple calculations show that marginally extending the scope of the LTU would create substantial net welfare gains by reducing resource costs of evasion in a magnitude much larger than the additional administrative costs required. This gap is due to the non-exhausted complementarity between monitoring effort and enforcement technology that leads to higher tax compliance induced by the threat of audit. Once the fixed-cost investment of setting up a tax enforcement technology has been made, the marginal cost of increasing human resources to exploit all the available information is relatively low and reduces the administrative costs of tax enforcement per taxpayer.

The existence of significant welfare gains at the margin also suggests the presence of sizable welfare gains in the interior of the LTU. As discussed in section 5.3, the LTU creates a tax compliance effect on large firms well above the threshold, implying significant increases of tax bases (in euros) and thus creating substantial welfare gains. Given the small magnitude of the local welfare costs and the low additional administrative costs of creating the LTU, the evidence suggests that the Spanish LTU has been effective at

[48] These figures correspond to 2007, the last year in our period of analysis. The average annual compensation of LTU tax auditors was lower in nominal terms in the earlier years, but we take the largest numbers to obtain a conservative estimate of the administrative cost of extending the LTU.

[49] Again, this is a conservative estimate because we are assuming that the marginal cost of including one additional taxpayer with reported revenue just below €6 million in the LTU census is equal to the *average* cost of monitoring taxpayers with reported revenue above that level.

[50] Official reports (AEAT, 1995-2007) show that tax auditors in the LTU perform intensive verifications on approximately 11% of LTU firms each year (i.e., roughly over 3,000 firms), mainly through electronic notifications from cross-checkings of information sources. As an example of the potential tax revenue gains of the policy, official statistics indicate that in 2004 the direct effect of the LTU monitoring through verifications yielded more than €800 million in additional tax revenue, to which we should add the indirect impact on other non-audited taxpayers' behavior due to the threat of audit effect.

increasing tax revenue capacity and also at raising welfare.[51]

Optimal Threshold. One natural question given the above analysis is what would be the optimal revenue level of the LTU threshold. We perform some back-of-the-envelope calculations here, which require making a number of strong assumptions. First, we assume that marginal resource costs ($\kappa_u \approx 0.059$), marginal administrative costs ($c_\phi \approx$ €333), and the percentage change in reported taxable income induced by the LTU's change in monitoring effort ($\frac{\Delta u}{\overline{y}} \approx 1\%$) are the same at all levels of reported revenue. Given these assumptions, we can easily calculate the revenue threshold at which the benefits of stricter tax enforcement would be exhausted at the point where $\Delta u \cdot \kappa_u = c_\phi$. Substituting-in the assumed values, we have $0.01\overline{y}^* \cdot 0.059 =$ €333, and therefore $\overline{y}^* =$ €564,400.

As an alternative, using the upper bound of the marginal resource costs (19%) yields a reported revenue threshold where welfare gains are exhausted at €175,200 (i.e. potential tax savings in CIT, PRT and VAT by marginal resource costs equal the marginal administrative costs of €333). The expansion of the LTU monitoring effort would thus require extending stricter monitoring to 197,000 and 383,000 additional firms, respectively.[52]

There is, of course, substantial uncertainty about the exact location of the optimal threshold, because it depends on strong assumptions and on extrapolation of our inherently local estimates. Nevertheless, these simple calculations are useful to illustrate the scope that governments have to extend their tax revenue capacity when they have already incurred the fixed cost of setting up a modern tax enforcement technology. Indeed, many LTUs designed to target large taxpayers may not be exploiting the economies of scale associated to extending stricter monitoring to smaller firms. On the other hand, notice that our calculations indicate that applying high monitoring effort to *all* small firms is not welfare-maximizing, as the estimated optimal thresholds exclude the vast majority of firms (80% or 60%, depending on the estimates chosen) from the LTU.[53] Finally, note that the insights raised in this analytical framework are applicable to the design of LTUs and other size-dependent enforcement policies around the world, adjusted to the specific

[51] According to official data (AEAT, 1995-2007), the additional administrative costs devoted to the Spanish LTU are on average €7.5-€10 million per year. Assuming a full evasion response of the bunchers (on average 250 firms per year) that on average misreport revenue by €300,000, implies that the local welfare costs are in the range of €4.42-€14.25 million, given the discussed bounds on the marginal resource costs of evasion. The observed tax base reporting regime shift above the threshold more than compensates for these losses.

[52] The calculations use firm-level data for 2007. Under the lower (upper) bound of marginal resource costs, this expansion would require hiring about 816 (1,595) tax auditors, with an expected annual cost of €65 (€128) million. This expansion would also require increasing the fixed-costs of the tax agency (e.g., buildings needed to locate new auditors) that are difficult to estimate, but could potentially be funded with the additional revenue raised by the LTU.

[53] This is consistent with the insights from the theoretical work by Dharmapala, Slemrod and Wilson (2011), who show that when there are high per-firm administrative costs of taxation and compliance costs for businesses, it may be optimal to exclude small firms from taxation.

conditions of each context of analysis.

7 Concluding Remarks

In this paper, we have investigated the effects of size-dependent tax enforcement on firms' tax compliance and the welfare implications of this type of policy. We first derive theoretical predictions on how firms respond to increases in the resources used by the tax authority to verify the transactions reported by firms. Then, we test the predictions of the model using quasi-experimental variation in monitoring effort provided by the Large Taxpayers Unit (LTU) in Spain.

The empirical results show that firms react to avoid being under stricter tax enforcement by reducing their reported revenue to bunch just below the LTU eligibility threshold. This response is heterogeneous across firms depending on the traceability of their transactions, suggesting that monitoring effort and information requirements are complementary to increase tax compliance. In particular, we find a larger bunching response in sectors that sell a large share of intermediate goods, where information trails are easier to verify by using more monitoring resources. Finally, we document that firms monitored by the LTU report larger tax bases, indicating that the policy is effective at reducing tax evasion. In contrast, small firms outside the LTU's radar have wide scope to misreport both labor and material expenditures to evade multiple taxes because of the low monitoring effort applied to them.

Incorporating these empirical findings into a welfare calculation, we conclude that devoting additional resources to extending stricter tax monitoring to smaller firms would generate net welfare gains at the margin. Back-of-the-envelope calculations of the optimal tax-enforcement threshold indicate that the optimal LTU threshold would be lower than the current one in Spain, but it would still leave a substantial fraction of small firms outside of the LTU. The general analytical framework presented here can be easily adapted to study the welfare consequences of LTUs and size-dependent tax enforcement policies more generally in other settings.

References

AEAT. 1995-2007. "Memorias de Administración Tributaria." Agencia Estatal de Administración Tributaria (Spanish Tax Authority).

Almunia, Miguel, and David Lopez-Rodriguez. 2014. "Heterogeneous Responses to Effective Tax Enforcement: Evidence from Spanish Firms." Working Paper 1419, Banco de España.

Bentolila, Samuel, Juan J. Dolado, and Juan F. Jimeno. 2012. "Reforming an insider-outsider labor market: the Spanish experience." *IZA Journal of European Labor Studies*, 1(4).

Best, Michael. 2013. "Salary Misreporting and the Role of Firms in Workers' Response to Taxes: Evidence from Pakistan." working paper, London School of Economics.

Best, Michael, and Henrik Kleven. 2016. "Housing Market Responses to Transaction Taxes: Evidence from Notches and Stimulus in the UK." *Review of Economic Studies*, forthcoming.

Best, Michael, Anne Brockmeyer, Henrik Kleven, Johannes Spinnewijn, and Mazhar Waseem. 2015. "Production vs Revenue Efficiency with Limited Tax Capacity: Theory and Evidence from Pakistan." *Journal of Political Economy*, 123: 1311–1355.

Bigio, Saki, and Eduardo Zilberman. 2011. "Optimal self-employment income tax enforcement." *Journal of Public Economics*, 95: 1021–1035.

Bird, Richard, and Pierre-Pascal Gendron. 2007. *The VAT in Developing and Transitional Countries.* Cambridge University Press.

Carrillo, Paul, Dina Pomeranz, and Monica Singhal. 2016. "Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement." *American Economic Journal: Applied Economics*, forthcoming.

Chetty, Raj. 2009*a*. "Is the Taxable Income Elasticity a Sufficient Statistic to Calculate Deadweight Loss? The Implications of Evasion and Avoidance." *American Economic Journal: Economic Policy*, 1(2): 31–52.

Chetty, Raj. 2009*b*. "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Analysis." *Annual Review of Economics*, 1: 451–488.

Chetty, Raj, John Friedman, Tore Olsen, and Luigi Pistaferri. 2011. "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records." *Quarterly Journal of Economics*, 126: 749–804.

Devereux, Michael, Li Liu, and Simon Loretz. 2014. "The Elasticity of Corporate Taxable Income: New Evidence from UK Tax Records." *American Economic Journal: Economic Policy*, 6(2): 19–53.

Dharmapala, Dhammika, Joel Slemrod, and John Douglas Wilson. 2011. "Tax policy and the missing middle: Optimal tax remittance with firm-level administrative costs." *Journal of Public Economics*, 95(9-10): 1036–1047.

Feldstein, Martin. 1999. "Tax Avoidance and The Deadweight Loss of Income Taxation." *Review of Economics and Statistics*, 81(4): 674–680.

Garicano, Luis, Claire Le Large, and John van Reenen. 2016. "Firm Size Distortions and the Productivity Distribution: Evidence from France." *American Economic Review*, 106(11): 3439–3479.

Gordon, Roger, and Wei Li. 2009. "Tax Structures in Developing Countries, Many Puzzles and a Possible Explanation." *Journal of Public Economics*, 93: 855–866.

Gorodnichenko, Yuriy, Jorge Martinez-Vazquez, and Klara Sabirianova Peter. 2009. "Myth and Reality of Flat Tax Reform: Micro Estimates of Tax Evasion Response and Welfare Effect in Russia." *Journal of Political Economy*, 117(3): 504–554.

Guner, Nezih, Gustavo Ventura, and Yi Xu. 2008. "Macroeconomic Implications of Size-Dependent Policies." *Review of Economic Dynamics*, 11: 721–744.

IEF. 2011. "Evolución del Sistema Fiscal Español: 1978-2010." Instituto de Estudios Fiscales.

IMF. 2002. "Improving Large Taxpayers' Compliance: A Review of Country Experience." International Monetary Fund, Washington DC. Team led by Katherine Baer.

IRS. 2012. "Tax Gap for Year 2006." Internal Revenue Service.

Kanbur, Ravi, and Michael Keen. 2014. "Thresholds, informality, and partitions of compliance." *International Tax and Public Finance Tax Public Finance*, 21: 536–559.

Keen, Michael, and Benjamin Lockwood. 2010. "The value added tax: Its causes and consequences." *Journal of Development Economics*, 92(2): 138–151.

Keen, Michael, and Jack Mintz. 2004. "The Optimal Threshold for a Value-Added Tax." *Journal of Public Economics*, 88: 559–576.

Kleven, Henrik. 2016. "Bunching." *Annual Review of Economics*, 8: 435–464.

Kleven, Henrik, and Mazhar Waseem. 2013. "Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan." *Quarterly Journal of Economics*, 128(2): 669–723.

Kleven, Henrik, Claus Kreiner, and Emmanuel Saez. 2016. "Why Can Modern Governments Tax So Much? An Agency Model of Firms as Fiscal Intermediaries." *Economica*, 83: 219–243.

Kleven, Henrik, Martin Knudsen, Claus Kreiner, Soren Pedersen, and Emmanuel Saez. 2011. "Unable or Unwilling to Cheat? Evidence from a Tax Audit Experiment in Denmark." *Econometrica*, 79(3): 651–692.

Kopczuk, Wojciech, and David J. Munroe. 2015. "Mansion Tax: The Effect of Transfer Taxes on the Residential Real Estate Market." *American Economic Journal: Economic Policy*, 7(2): 214–57.

Kopczuk, Wojciech, and Joel Slemrod. 2006. "Putting Firms into Optimal Tax Theory." *American Economic Review, Papers and Proceedings*, 96(2): 130–134.

Kumler, Todd, Eric Verhoogen, and Judith Frias. 2015. "Enlisting Employees in Improving Payroll-Tax Compliance: Evidence from Mexico." working paper, Columbia University.

Liu, Li, and Benjamin Lockwood. 2015. "VAT Notches." CESifo Working Paper No. 5371.

Lucas, Robert. 1978. "On the Size Distribution of Business Firms." *The Bell Journal of Economics*, 9(2): 508–523.

Naritomi, Joana. 2015. "Consumers as Tax Auditors." working paper, London School of Economics.

OECD. 2011. *Tax Administration in OECD and Selected non-OECD Countries: Comparative Information Series.* Organisation for Economic Cooperation and Development.

Onji, Kazuki. 2009. "The Response of Firms to Eligibility Thresholds: Evidence from the Japanese Value-Added Tax." *Journal of Public Economics*, 93: 766–775.

Paula, Áureo De, and José Scheinkman. 2010. "Value-Added Taxes, Chain Effects, and Informality." *American Economic Journal: Macroeconomics*, 2: 195–221.

Pomeranz, Dina. 2015. "No Taxation Without Information: Deterrence and Self-Enforcement in the Value-Added Tax." *American Economic Review*, 105(8): 2539–69.

Restuccia, Diego, and Richard Rogerson. 2008. "Policy Distortions and Aggregate Productivity with Heterogeneous Establishments." *Review of Economic Dynamics*, 11: 707–720.

Saez, Emmanuel. 2010. "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy*, 2(3): 180–212.

Slemrod, Joel. 2010. "Buenas Notches: Lines and Notches in Tax System Design." working paper, University of Michigan.

Slemrod, Joel, and Christian Gillitzer. 2014. *Tax Systems.* Cambridge:MIT Press.

Slemrod, Joel, and Shlomo Yitzhaki. 1987. "The Optimal Size of a Tax Collection Agency." *The Scandinavian Journal of Economics*, 89(2): 183–192.

Slemrod, Joel, Brett Collins, Jeffrey Hoopes, Daniel Reck, and Michael Sebastiani. 2015. "Does Credit-Card Information Reporting Improve Small-Business Tax Compliance?" working paper, University of Michigan.

Slemrod, Joel, Marsha Blumenthal, and Charles Christian. 2001. "Taxpayer Response to an Increased Probability of Audit: Evidence from a Controlled Experiment in Minnesota." *Journal of Public Economics*, 79: 455–483.

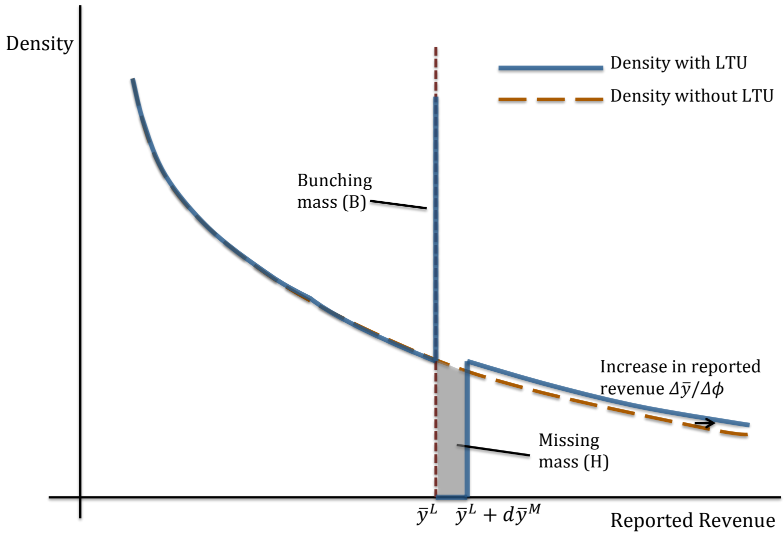
Tribunal de Cuentas. 2008. "Informe de Fiscalización de las Unidades de Gestión de Grandes Empresas de la AEAT." Tribunal de Cuentas.

Yaniv, Gideon. 1988. "Withholding and Non-Withheld Tax Evasion." *Journal of Public Economics*, 35: 183–204.

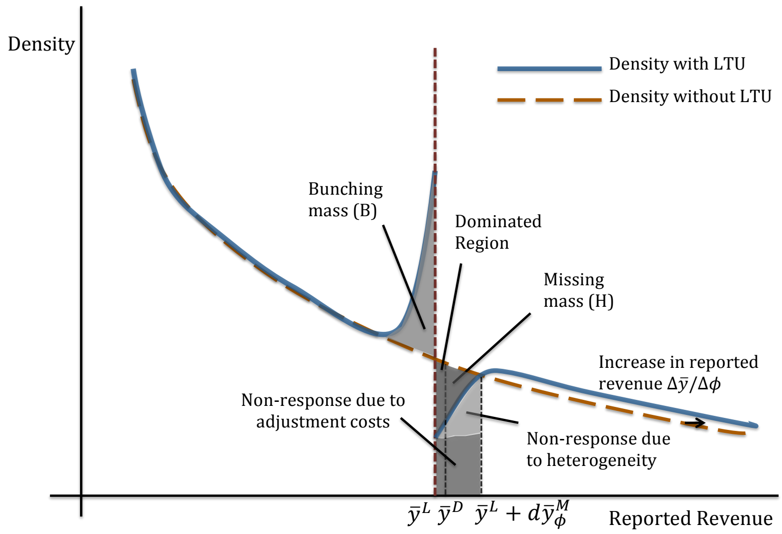
Figures

Figure 1: Theoretical Distribution of Reported Revenue

(a) Baseline Model

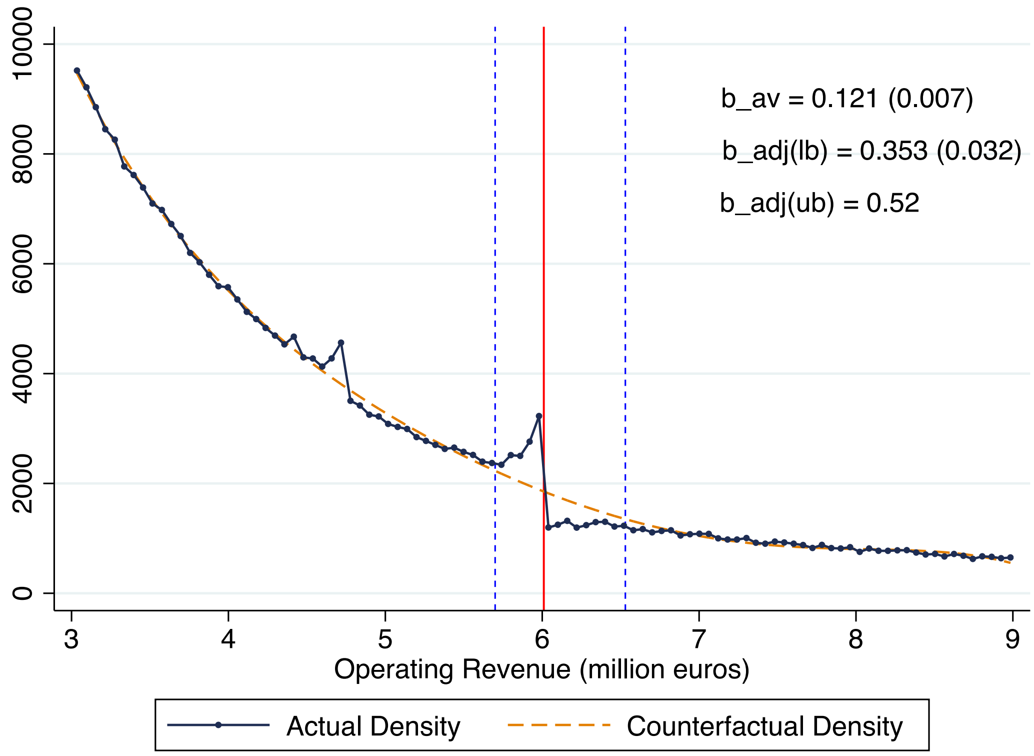


(b) Model with Heterogeneous Firms



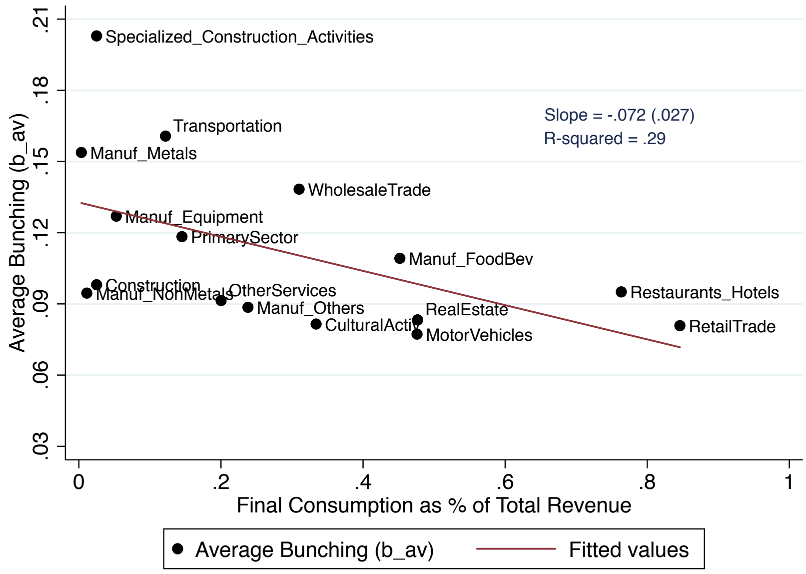
Notes: this figure depicts the theoretical revenue distribution before and after the introduction of the Large Taxpayers Unit (LTU). Without the LTU, all firms face the same monitoring effort and the distribution of revenue is smoothly decreasing as depicted by the dashed (brown) line. When the LTU is introduced, firms reporting revenue above y^L face a higher enforcement intensity. A group of firms in an interval above y^L respond to the new policy by underreporting more of their revenue to report exactly $\bar{y} = y^L$. This generates a spike at the threshold (with excess mass B), and an area of missing mass (H) to the right of the threshold, as depicted by the solid (blue) line. Panel (a) corresponds to the baseline model, where monitoring effort and resource costs of evasion are the same for all firms, so all firms with the same productivity draw respond identically to fiscal incentives. Thus, no firm locates in the interval of length $d\bar{y}^M$ to the right of the LTU threshold. Panel (b) depicts the equilibrium in the model with heterogeneity, where a fraction of firms does not respond to the incentives due to different changes in monitoring effort at the threshold or due to other adjustment costs.

Figure 2: Operating Revenue Distribution (1995-2007): Observed vs. Counterfactual



Notes: this graph shows the reported distribution of revenue (dots connected by solid blue line) and the estimated counterfactual (orange dashed curve), using pooled data for the period 1995-2007. The vertical red line indicates the Large Taxpayers Unit (LTU) threshold (€6 million). The vertical dashed blue lines indicate the bounds of the excluded region (y_{lb} and y_{ub}) chosen for the estimation of the counterfactual. To determine the value of y_{ub}, we fit a 5th-degree polynomial to the true density in multiple iterations, starting with $y_{ub} \approx y^L$ and increasing the value in small steps until we reach a point where the bunching mass (B) equals the missing mass (H), so that the integration constraint is satisfied. The average bunching parameter (b_{av}) estimates the adjustment in reported revenue for the average firm above the threshold, while (b_{adj}) estimates the adjustment for the marginal bunching firm, accounting for the existence of resource costs of evasion that prevent some firms from responding to the notch. The bins are €60,101 wide, delimited such that no bin contains data both to the left and to the right of the relevant policy thresholds. The total number of observations is 285,570.

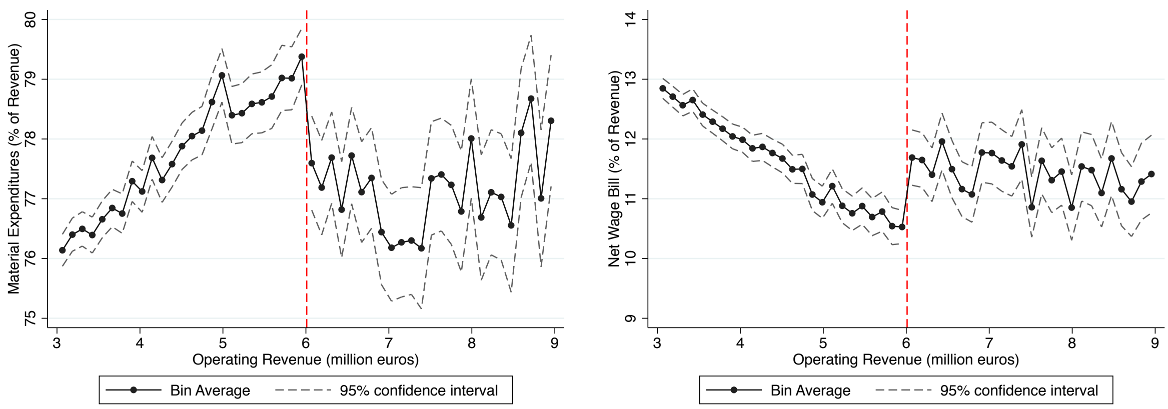
Figure 3: Average Bunching by Share of Final Consumer Sales



Notes: the bunching measure $\widehat{b}_{av}$ is calculated for each sector as explained in section 3 in the main text. Final consumption as a share of total sales in each sector is calculated using the year 2000 input-output tables for the Spanish economy, published by the National Statistics Institute (INE). More details about this data source are provided in online appendix D. The figure shows a negative relationship between the average bunching response (b_{av}) and the percentage of sales made to final consumers in each sector. The slope coefficient reported is obtained by estimating a simple linear regression of the bunching estimates on final consumption share in each sector, using robust standard errors. See Table 3 for regressions with additional controls.

Figure 4: Reported Input Expenditures

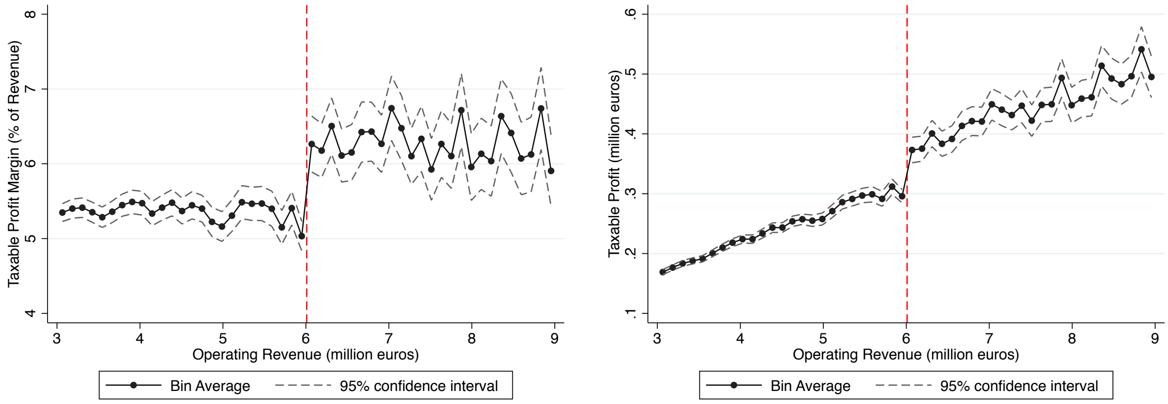
(a) Material Expenditures (b) Labor Expenditures



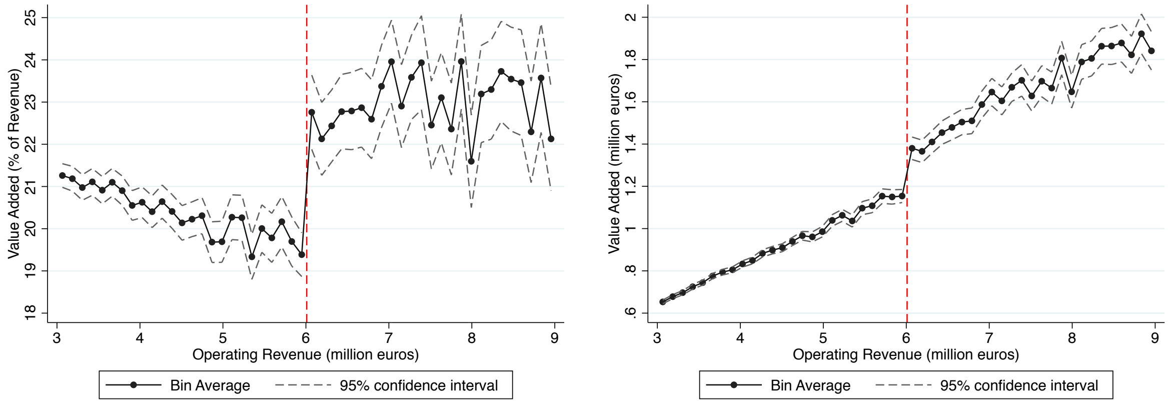
Notes: these graphs show the average ratio of input expenditures as a proportion of revenue (vertical axes) against reported revenue (horizontal axes), for the period 1995-2007. "Material expenditures" includes all the intermediate inputs used by the firm for production. "Labor expenditures" is the total wage bill of the firm, excluding employee-contributed payroll taxes (social security contributions). The dashed (red) vertical line indicates the LTU threshold. The dotted lines denote bin averages and the gray dashed lines show 95% confidence intervals for each bin average. To avoid the spurious effect of extreme values, we winsorize observations in the top and bottom 1% of the outcome variable, meaning that we set those values equal to the first and 99th percentile. We do this for each €1-million interval in the range $\overline{y} \in (3, 9)$ million. (Bin width= €120,202).

Figure 5: Reported Tax Bases

(a) Taxable Profits: Tax Base of the Corporate Income Tax



(b) Value Added: Tax Base of the VAT



Notes: these graphs show the average tax bases reported by firms in the revenue interval $\overline{y} \in (3, 9)$ million. The black dotted lines denote bin averages and the gray dashed lines show 95% confidence intervals for each bin average. Panel (a) shows the reported average taxable profits as a percentage of total revenue (left) and in million euros (right). Since taxable profits are not directly reported in the data, we calculate them out using the corporate income tax liability reported by firms in their financial schedule and apply the tax schedule. Note that this calculation can only be made for firms with a positive tax liability, which in this period were more than 80% of all firms in the sample. Panel (b) shows average reported value added as a percentage of total revenue (left) and in million euros (right). This is the tax base of the VAT. Note that the base of the payroll tax is the wage bill, shown in Figure 4b. To avoid the spurious effect of extreme values, we winsorize observations in the top and bottom 1% of the outcome variable, meaning that we set those values equal to the first and 99th percentile. We do this for each €1-million interval in the range $\overline{y} \in (3, 9)$ million. (Bin width= €120,202).

Tables

Table 1: Bunching Estimations by Year

	Bunching Estimators		# of Firms		Excl. Interval		Obs.
	b_{av}	b_{adj}	B	H	y_{lb}	y_{ub}	N
Pooled data							
1995-2007	0.121	0.382	3191	2965	5.70	6.53	285,570
	(0.007)***	(0.036)***					
Annual data							
1995	0.089	0.257	82	77	5.70	6.56	10,207
	(0.019)***	(0.079)***					
1996	0.099	0.395	98	91	5.70	6.38	11,014
	(0.014)***	(0.111)***					
1997	0.171	0.515	186	183	5.70	6.59	12,516
	(0.022)***	(0.118)***					
1998	0.163	0.351	198	185	5.70	6.53	14,310
	(0.024)***	(0.077)***					
1999	0.172	0.416	218	203	5.70	6.68	15,835
	(0.026)***	(0.097)***					
2000	0.194	0.660	273	251	5.70	6.80	17,923
	(0.025)***	(0.166)***					
2001	0.130	0.350	229	214	5.70	6.59	21,030
	(0.018)***	(0.073)***					
2002	0.078	0.410	168	151	5.70	6.38	23,977
	(0.010)***	(0.109)***					
2003	0.102	0.348	249	254	5.70	6.41	26,759
	(0.014)***	(0.080)***					
2004	0.147	0.550	382	356	5.70	6.86	29,779
	(0.018)***	(0.132)***					
2005	0.110	0.321	382	376	5.70	6.47	33,342
	(0.010)***	(0.045)***					
2006	0.095	0.298	373	342	5.70	6.35	35,788
	(0.009)***	(0.044)***					
2007	0.099	0.305	336	310	5.70	6.50	33,090
	(0.011)***	(0.053)***					

Notes: b_{av} is the average bunching response and b_{adj} is the marginal buncher's response taking into account resource costs of evasion (both measured in million euros). Bootstrapped standard errors are shown below each estimate in parenthesis. B is the number of firms above the counterfactual density of revenue in the range $y \in (y_{lb}, y^L)$, where y is revenue, y_{lb} is the lower bound of the excluded region (used to construct the counterfactual) and y^L is the LTU threshold of €6 million. H is the missing number of firms below the counterfactual density in the range $y \in (y^L, y_{ub})$, where y_{ub} is the upper bound of the excluded region. The upper and lower bounds of the excluded interval, (y_{lb}, y_{ub}) are also reported. For all years, the counterfactual density is estimated using a 5th-degree polynomial. Finally, N is the number of observations included in the estimations, i.e. the number of firms with revenue $y \in$(€3.01,€9.01) million in each year. Significance levels: *** = 1%, ** = 5%, and * = 10%.

Table 2: Bunching Estimations, by Sector of Activity

	Bunching Estimators		# of Firms		Excl. Interval		Obs.
	b_{av}	b_{adj}	B	H	y_{lb}	y_{ub}	N
Primary Sector	0.118	1.301	40	37	5.70	6.53	4,120
	(0.034)***	(21.917)					
Manuf. Food and Beverages	0.109	0.533	105	108	5.70	6.89	10,100
	(0.023)***	(0.320)**					
Manuf. Non Metals	0.095	0.245	97	89	5.70	6.29	9,415
	(0.014)***	(0.050)***					
Manuf. Metals	0.154	0.528	115	103	5.70	6.47	8,262
	(0.023)***	(0.152)***					
Manuf. Equipment	0.127	0.300	73	67	5.70	6.68	6,137
	(0.029)***	(0.107)***					
Manuf. Others	0.089	0.291	142	128	5.70	6.53	18,818
	(0.016)***	(0.079)***					
Construction of Buildings	0.098	0.478	272	251	5.70	6.62	29,070
	(0.011)***	(0.105)***					
Specialized Constr. Activ.	0.203	0.496	180	170	5.70	6.74	11,838
	(0.025)***	(0.097)***					
Motor Vehicles	0.077	0.361	118	107	5.70	6.41	13,118
	(0.009)***	(0.077)***					
Wholesale (exc. Motor V.)	0.138	0.342	986	939	5.70	6.47	72,865
	(0.008)***	(0.029)***					
Transportation	0.161	0.485	177	161	5.70	6.80	13,025
	(0.023)***	(0.124)***					
Retail Trade	0.081	0.283	155	146	5.70	6.29	22,053
	(0.011)***	(0.061)***					
Restaurants and Hotels	0.095	0.470	64	59	5.70	6.74	7,868
	(0.028)***	(0.389)					
Cultural Activities	0.082	0.393	38	37	5.70	6.38	5,397
	(0.027)***	(0.439)					
Real Estate	0.083	0.952	38	36	5.70	6.38	4,829
	(0.022)***	(5.939)					
Other Services	0.091	0.326	146	147	5.70	6.53	16,799
	(0.012)***	(0.071)***					

Notes: b_{av} is the average bunching response and b_{adj} is the marginal buncher's response taking into account resource costs of evasion (both measured in million euros). Bootstrapped standard errors are shown below each estimate in parenthesis. B is the number of firms above the counterfactual density of revenue in the range $y \in (y_{lb}, y^L)$, where y is revenue, y_{lb} is the lower bound of the excluded region (used to construct the counterfactual) and y^L is the LTU threshold of €6 million. H is the missing number of firms below the counterfactual density in the range $y \in (y^L, y_{ub})$, where y_{ub} is the upper bound of the excluded region. The upper and lower bounds of the excluded interval, (y_{lb}, y_{ub}) are also reported. For all years, the counterfactual density is estimated using a 5th-degree polynomial. Finally, N is the number of observations included in the estimations, i.e. the number of firms with revenue $y \in$(€3.01,€9.01) million in each year. Significance levels: *** = 1%, ** = 5%, and * = 10%.

Table 3: Determinants of Average Bunching Response

(a) Ordinary Least Squares Regressions

	Dependent variable: Average Bunching (b_{av})					
	(1)	(2)	(3)	(4)	(5)	(6)
Share of Final Consumer Sales	-0.072	-0.071	-0.070	-0.053	-0.068	-0.054
	(0.027)**	(0.026)**	(0.026)**	(0.022)**	(0.024)**	(0.023)**
Number of Employees (median)		0.000			0.000	0.000
		(0.001)			(0.001)	(0.001)
Tangible Fixed Assets (median)			-0.023		-0.023	0.001
			(0.025)		(0.026)	(0.037)
Share of Non-Bunchers (α)				-0.124		-0.116
				(0.070)		(0.107)
Constant	0.133	0.122	0.143	0.194	0.132	0.192
	(0.014)***	(0.017)***	(0.023)***	(0.041)***	(0.026)***	(0.051)***
Observations	16	16	16	16	16	16
R-squared	0.286	0.301	0.319	0.400	0.336	0.401

Notes: All regressions are estimated by OLS. The unit of observation is the sector of activity (for details on the definition of the 16 sectors, see online appendix D). The median number of employees in each sector refers to full-time-equivalent (FTE). Tangible fixed assets are measured in million euros. Robust standard errors reported in parentheses. Significance levels: *** = 1%, ** = 5%, and * = 10%.

(b) Weighted Least Squares Regressions

	Dependent variable: Average Bunching (b_{av})					
	(1)	(2)	(3)	(4)	(5)	(6)
Share of Final Consumer Sales	-0.049	-0.051	-0.050	-0.035	-0.051	-0.038
	(0.014)***	(0.015)***	(0.014)***	(0.015)**	(0.015)***	(0.015)**
Number of Employees (median)		-0.000			-0.000	-0.001
		(0.000)			(0.000)	(0.000)
Tangible Fixed Assets (median)			-0.021		-0.019	0.008
			(0.015)		(0.016)	(0.017)
Share of Non-Bunchers (α)				-0.188		-0.191
				(0.043)***		(0.044)***
Constant	0.121	0.127	0.128	0.225	0.130	0.244
	(0.006)***	(0.010)***	(0.008)***	(0.025)***	(0.011)***	(0.029)***
Observations	16	16	16	16	16	16

Notes: All regressions are estimated by Weighted Least Squares. The unit of observation is the sector of activity (for details on the definition of the 16 sectors, see online appendix D). The weights are the inverse of the variance of the bunching estimates. The median number of employees in each sector refers to full-time-equivalent (FTE). Tangible fixed assets are measured in million euros. Standard errors reported in parentheses. Significance levels: *** = 1%, ** = 5%, and * = 10%.

Online Appendix
Not Intended for Publication

"Under the Radar: The Effects of Monitoring Firms on Tax Compliance"

Miguel Almunia (University of Warwick)
David Lopez-Rodriguez (Banco de España)

A Corporate Evasion with Multiple Taxes

Consider a firm that produces good y combining material acquisitions m and labor n, which are tax-deductible expenditures[54] in the corporate income tax, and non-deductible inputs z according to the production function $y = \psi f(m, n, z)$, where ψ is a productivity parameter and $f(\cdot, \cdot, \cdot)$ is strictly continuous, increasing and concave in inputs use. Firm hires in competitive markets materials at unit cost c, labor at wage rate w, and non-deductible expenditures at unit cost q, and sell their output at the market price p, which is normalized to unity.

Suppose the existence of value added tax (VAT) by the credit method in which firms charge a flat tax rate t^{vat} on their sales y and receive a credit for the monetary value of their material expenditures $e \equiv cm$. Firms must transfer to the tax authority the difference between charged and deductible VAT, that is $t^{vat} \cdot P^{vat}$ with $P^{vat} = y - e$. Government also levies linear payroll taxes on the wage bill $P^{ss} = l \equiv wn$, charging t^{ss_1} on account of employers, that are tax deductible in the corporate income tax, and t^{ss_2} on account of employers. We assume that both payroll taxes are fully born by firms. Firms also consider the regulatory costs associated to hiring labor captured by a convex cost function in the reported wage bill, $\gamma(l)$.[55] Finally, the income generated by the firm is taxed with a proportional rate t^{cit} on taxable profits $P^{cit} = y - e - l \cdot (1 + t^{ss_1})$, so firm's net-of-tax income with truthful reporting is given by $\Pi = (1-t^{cit})P^{cit} - P^{ss}t^{ss_2} - qz - \gamma(l)$.

Suppose that the tax authority is not able to monitor all transactions in the economy creating incentives for firms to evade taxes by misreporting their tax bases. Consider that an evader firm could underreport the monetary value of their revenue by an amount $u^y \equiv y - \overline{y} \geq 0$, where $\overline{y}$ denotes reported revenue, to reduce taxable corporate income and to appropriate tax revenue from the VAT. Firm may also attempt to inflate the value of their material acquisitions, given by $u^e \equiv \overline{e} - e \geq 0$, where $\overline{e}$ denotes reported expenditures, to claim larger tax credits in both corporate income tax and the VAT. Firms may have incentives to hide a share of their wage bill by an amount $u^l \equiv l - \overline{l} \geq 0$, where $\overline{l}$ denotes reported labor expenditures, to evade payroll taxes and save regulatory costs of hiring labor. Given these potential evasion channels, firm's reported tax bases in the corporate income tax, payroll taxes and the VAT are given, respectively, by

$$\overline{P^{cit}} = [(y - u^y) - (e + u^e) - (l - u^l) \cdot (1 + t^{ss_1})], \tag{18}$$

[54] We make the distinction between this two tax-deductible inputs because the dataset in our empirical application includes accurate measures of firms' total expenditures on material acquisitions and labor wage bill.

[55] The assumptions on the incidence of payroll taxes on account of employers and employees, and the existence of regulatory costs associated to hiring workers seems particularly appropriate for the Spanish case. As an example, Bentolila, Dolado and Jimeno (2012) discuss the costs and rigidities imposed on Spanish firms by multiple regulations in labor markets.

$$\overline{P^{ss}} = (l - u^l), \tag{19}$$

$$\text{and} \quad \overline{P^{vat}} = [(y - u^y) - (e + u^e)]. \tag{20}$$

Evasion behavior is costly because it requires, for instance, collusion between the firm and its trading partners and employees; the creation of parallel accounting books and payment systems in cash; or it can imply forego business opportunities. We introduce these resource costs of evasion by a reduced form $\kappa(u^y, u^e, u^l)$ that is an increasing, convex and separable function in each of its arguments. The tax authority detects evasion with probability $\delta = \phi h(u^y, u^e, u^l)$, where ϕ is the monitoring effort parameter and the enforcement technology $h(\cdot)$ is a continuous, convex and separable function in each evasion channel. Whenever misreporting is detected, the firm is compelled to pay back the evaded tax plus a proportional penalty rate θ that, for simplicity, is assumed homogeneous for all channels of evasion.

The expected profit of the firm net of corporate and payroll taxes, and augmented by the expected appropriation of VAT revenue, is given by

$$\begin{aligned} E\Pi \;=\; & (1 - t^{cit})[\psi f(m, n, z) - e - l(1 + t^{ss_1})] - qz + tr \cdot [u^y + u^e - u^l(1 + t^{ss_1})] \\ & + (t^{ss_1} + t^{ss_2}) \cdot ru^l + t^{vat} \cdot r[u^y + u^e] - \kappa(u^y, u^e, u^l) - \gamma(\bar{l}), \end{aligned} \tag{21}$$

where $r \equiv [1 - \phi h(u^y, u^e, u^l)(1 + \theta)]$ is the expected rate of return of 1 euro evaded. Firms make production and reporting decisions in order to maximize their expected profit such that an interior optimum for firms real and evasion decisions satisfies the system of first-order conditions given by

$$\psi f_m(m^*, n^*, z^*) = c \tag{22}$$

$$\psi f_n(m^*, n^*, z^*) = w \left[1 + t^{ss_1} + \frac{t^{ss_2} + \gamma_{\bar{l}}(\bar{l}^*)}{(1 - t^{cit})} \right] \tag{23}$$

$$\psi f_z(m^*, n^*, z^*) = q/(1 - t) \tag{24}$$

$$\left[t^{cit} + t^{vat} \right] \cdot r = \kappa_{u^y}(u^{y*}) + (1 + \theta)\phi h_{u^y}(u^{y*}) \cdot \widehat{T} \tag{25}$$

$$\left[t^{cit} + t^{vat} \right] \cdot r = \kappa_{u^e}(u^{e*}) + (1 + \theta)\phi h_{u^e}(u^{e*}) \cdot \widehat{T} \tag{26}$$

$$\left[(t^{ss_1} + t^{ss_2}) - t^{cit}(1 + t^{ss_1}) \right] \cdot r + \gamma_{\bar{l}}(\bar{l}^*) = \kappa_{u^l}(u^{l*}) + (1 + \theta)\phi h_{u^l}(u^{l*}) \cdot \widehat{T} \tag{27}$$

where $\widehat{T} \equiv [t^{vat} \cdot (u^{y*} + u^{e*}) + (t^{ss_1} + t^{ss_2}) \cdot u^{l*} + t^{cit} \cdot (u^{y*} + u^{e*} - u^{l*}(1 + t^{ss_1}))]$ is the total evaded taxes by the multiple misreporting channels. The system of optimal conditions shows that positive tax rates on corporate income distort inputs demand decisions reducing revenue from potential production at zero tax rates. These conditions also indicate that the existence of both payroll taxes and labor regulatory costs create distortions increasing the marginal cost of hiring employees and thus reducing labor demand.

The optimal evasion conditions for each misreporting channel predict that firm evades taxes to the point where the marginal expected return of misreporting transactions is equal to the expected costs associated to tax evasion. The latter is the result of the marginal resource costs born in each misreporting channel plus the deterrence effect created by tax enforcement that results from the interaction between monitoring effort and the enforcement technology. The systematic matching of tax returns from multiple taxpayers implies that a marginal unit of misreporting in one channel increases the chances of being detected, and thus paying back the total amount evaded, in multiple channels.

The expected returns of misreporting revenue and expenditures are positively related with the tax rates. The larger the tax rates on both the VAT and the corporate income tax are, the higher the incentives to hide revenue and inflate material acquisitions to reduce those tax bases. Notice that when firms have scope to misreport their transactions they do not act as fiscal intermediaries, that just transfer collected VAT to the tax agency, but instead firms have incentives to appropriate a share of VAT revenue. Finally, the optimal condition for hidden labor bill indicates that firms could have incentives to misreport it when the marginal savings in payroll taxes and regulatory costs were larger than the foregoing tax credits in corporate income tax due misreporting of labor costs. Overall, the model identifies two channels that create positive returns for labor misreporting: i) the existence of a significant gap between payroll taxes and corporate tax rates; and ii) the presence of large regulatory costs associated to hiring workers.

B Anatomy of the LTU Response: Input Ratios and Tax Bases

Consider the model with heterogeneous monitoring effort and resource costs presented in subsection 2.2. Before the introduction of a LTU, the system of optimal conditions indicates that the demand of tax-deductible inputs (e.g. materials and labor) is smoothly increasing in productivity, $dm/d\psi > 0$ and $dn/d\psi > 0$. Hence, the reported ratios of input expenditures over revenue, $\overline{M} \equiv cm/\overline{y}$ and $\overline{L} \equiv wn/\overline{y}$, are continuous in ψ over the range $[\underline{\psi}, \overline{\psi}]$. This implies that in the neighborhood of y^L defined by the small interval $(y', y^L + d\overline{y}^M_\phi)$ the average reported ratios of inputs expenditures over revenue are *almost* *equal*, that is,

$$\frac{\int_{y'}^{y^L}\overline{M}\cdot g_0(\overline{y})d\overline{y}}{\int_{y'}^{y^L}g_0(\overline{y})d\overline{y}}\cong\frac{\int_{y^L}^{y^L+d\overline{y}_\phi^M}\overline{M}\cdot g_0(\overline{y})d\overline{y}}{\int_{y^L}^{y^L+d\overline{y}_\phi^M}g_0(\overline{y})d\overline{y}}\quad\text{and}\quad\frac{\int_{y'}^{y^L}\overline{L}\cdot g_0(\overline{y})d\overline{y}}{\int_{y'}^{y^L}g_0(\overline{y})d\overline{y}}\cong\frac{\int_{y^L}^{y^L+d\overline{y}_\phi^M}\overline{L}\cdot g_0(\overline{y})d\overline{y}}{\int_{y^L}^{y^L+d\overline{y}_\phi^M}g_0(\overline{y})d\overline{y}}. \tag{28}$$

Suppose the LTU is introduced. In the presence of heterogeneity in both monitoring effort effectiveness and resource costs, there is a subset of firms in the pre-LTU density interval $(y^L, y^L + d\overline{y}_\phi^M)$, the *non-optimizers*, with prohibitive resource costs to respond reducing reported revenue. In contrast, the complementary subset of firms also located in that interval in the pre-LTU situation, the *bunchers*, reduce their reported revenue to stay below the threshold because that results in larger expected profits, i.e. $E\Pi_0(m, n, z, u \mid \psi^M, \phi_0) > E\Pi_1(m, n, z, u \mid \psi^M, \phi_1)$. Considering that, due to frictions, the bunchers locate along the interval(y', y^L), the model provides different predictions on the expected average reported ratios of input expenditures over revenue around the LTU threshold. These predictions depend on whether bunchers' reaction is due to real (i.e. reduction of production) or evasion (i.e. increase of concealed revenue) responses to the enforcement threshold.[56]

Real Response. Bunchers can avoid the threshold lowering their production, and thus their inputs demand, without bearing additional resource costs of evasion. This reaction implies that in the interval (y', y^L) below the threshold there are firms with $\psi \in [\psi', \psi^L]$ that hire more inputs than bunchers with $\psi \in [\psi^L, \psi^M]$. This causes that both average reported ratios of expenditures over revenue are not continuous at the threshold y^L. Indeed, the real reaction of the bunchers to the LTU results in i) a downward trend of both ratios in the interval (y', y^L); and ii) a discrete upward jump of these ratios at the threshold such that

$$\frac{\int_{y'}^{y^L}\overline{M}\cdot g_1(\overline{y})d\overline{y}}{\int_{y'}^{y^L}g_1\left(\overline{y}\right)d\overline{y}}<\frac{\int_{y^L}^{y^L+d\overline{y}_\phi^M}\overline{M}\cdot g_1(\overline{y})d\overline{y}}{\int_{y^L}^{y^L+d\overline{y}_\phi^M}g_1(\overline{y})d\overline{y}}\quad\text{and}\quad\frac{\int_{y'}^{y^L}\overline{L}\cdot g_1(\overline{y})d\overline{y}}{\int_{y'}^{y^L}g_1\left(\overline{y}\right)d\overline{y}}<\frac{\int_{y^L}^{y^L+d\overline{y}_\phi^M}\overline{L}\cdot g_1(\overline{y})d\overline{y}}{\int_{y^L}^{y^L+d\overline{y}_\phi^M}g_1(\overline{y})d\overline{y}}. \tag{29}$$

Evasion Response. Bunchers can avoid the threshold increasing their concealed revenue, and thus without modifying their inputs demand, paying resource costs of additional evasion. This response implies that in the interval (y', y^L) below the threshold firms with $\psi \in [\psi', \psi^L]$ hire lower inputs than bunchers with $\psi \in [\psi^L, \psi^M]$. The evasion response of the bunchers thus creates i) an upward trend of the average reported ratios

[56] The bunchers' reaction to the LTU threshold could be a combination of both potential responses, real and evasion, but we discuss the two polar responses for analytical simplicity. This simplification provides predictions on the expected average patterns of input ratios around the threshold when the reaction is dominated by either the real or the evasion channel.

of expenditures in the interval (y', y^L); and ii) a discontinuous downward jump of these ratios at the threshold such that

$$\frac{\int_{y'}^{y^L} \overline{M} \cdot g_1(\overline{y}) d\overline{y}}{\int_{y'}^{y^L} g_1(\overline{y}) d\overline{y}} > \frac{\int_{y^L}^{y^L + d\overline{y}_\phi^M} \overline{M} \cdot g_1(\overline{y}) d\overline{y}}{\int_{y^L}^{y^L + d\overline{y}_\phi^M} g_1(\overline{y}) d\overline{y}} \quad \text{and} \quad \frac{\int_{y'}^{y^L} \overline{L} \cdot g_1(\overline{y}) d\overline{y}}{\int_{y'}^{y^L} g_1(\overline{y}) d\overline{y}} > \frac{\int_{y^L}^{y^L + d\overline{y}_\phi^M} \overline{L} \cdot g_1(\overline{y}) d\overline{y}}{\int_{y^L}^{y^L + d\overline{y}_\phi^M} g_1(\overline{y}) d\overline{y}} . \tag{30}$$

Evasion Response with Inputs Misreporting. Considering the extended model with inputs misreporting and multiple taxes presented above (Appendix B), evader firms have incentives to i) inflate their material acquisitions in an amount u^e to claim larger tax credits in both the VAT and the CIT; and ii) hide part of their wage bill, u^l, to save both payroll taxes and the regulatory costs of labor. The optimal amount of evasion in each expenditure channel is heterogeneous among firms because it depends negatively on the effectiveness of monitoring effort and the resource cost of evasion. Firms thus have larger incentives to avoid stricter tax enforcement when their expenditures misreporting is higher in the pre-LTU situation (i.e. larger expected profits of bunching). Bunchers that react to avoid the LTU increasing their concealed revenue therefore also report a higher (lower) proportion of materials (labor) to evade taxes than firms with $\psi \in [\psi', \psi^L]$ also located in the interval (y', y^L). Define the ratios of reported inputs expenditures over revenue with inputs misreporting by $\overline{M} \equiv (cm + u^e)/\overline{y}$ and $\overline{L} \equiv (wn - u^l)/\overline{y}$, respectively. The evasion response of bunchers that also misreport expenditures in a larger proportion creates i) an upward (downward) trend of the average ratio of materials (labor) in the interval (y', y^L); and ii) a downward (upward) jump of the materials (labor) ratio at the threshold such that

$$\frac{\int_{y'}^{y^L} \overline{M} \cdot g_1(\overline{y}) d\overline{y}}{\int_{y'}^{y^L} g_1(\overline{y}) d\overline{y}} >> \frac{\int_{y^L}^{y^L + d\overline{y}_\phi^M} \overline{M} \cdot g_1(\overline{y}) d\overline{y}}{\int_{y^L}^{y^L + d\overline{y}_\phi^M} g_1(\overline{y}) d\overline{y}} \quad \text{and} \quad \frac{\int_{y'}^{y^L} \overline{L} \cdot g_1(\overline{y}) d\overline{y}}{\int_{y'}^{y^L} g_1(\overline{y}) d\overline{y}} < \frac{\int_{y^L}^{y^L + d\overline{y}_\phi^M} \overline{L} \cdot g_1(\overline{y}) d\overline{y}}{\int_{y^L}^{y^L + d\overline{y}_\phi^M} g_1(\overline{y}) d\overline{y}} . \tag{31}$$

LTU Effectiveness: Tax Bases. High productivity firms that are in the interior of the LTU have lower scope to evade taxes when the LTU is effective ($d\phi > 0$). The extended model predicts that these firms reduce concealed outcome, u^y, inflated materials, u^e, and hidden wage bill, u^l, in a magnitude that depends on the effectiveness of the LTU. An effective LTU thus raises the reported tax bases of the corporate income tax, $\overline{P^{cit}}$, the payroll tax, $\overline{P^{ss}}$, and the value-added tax, $\overline{P^{vat}}$, with a break at the LTU threshold creating i) a downward trend of the average tax bases in the interval (y', y^L) due to bunchers' misreporting; and ii) an upward parallel shift of tax bases for high productivity firms such that

$$\frac{\int_{y'}^{y^L} \overline{P^{cit}} \cdot g_1(\overline{y})d\overline{y}}{\int_{y'}^{y^L} g_1(\overline{y})d\overline{y}} << \frac{\int_{y^L}^{y^L+d\overline{y}_{\phi}^M} \overline{P^{cit}} \cdot g_1(\overline{y})d\overline{y}}{\int_{y^L}^{y^L+d\overline{y}_{\phi}^M} g_1(\overline{y})d\overline{y}} \quad , \tag{32}$$

$$\frac{\int_{y'}^{y^L} \overline{P^{ss}} \cdot g_1(\overline{y})d\overline{y}}{\int_{y'}^{y^L} g_1(\overline{y})d\overline{y}} << \frac{\int_{y^L}^{y^L+d\overline{y}_{\phi}^M} \overline{P^{ss}} \cdot g_1(\overline{y})d\overline{y}}{\int_{y^L}^{y^L+d\overline{y}_{\phi}^M} g_1(\overline{y})d\overline{y}} \quad , \tag{33}$$

$$\frac{\int_{y'}^{y^L} \overline{P^{vat}} \cdot g_1(\overline{y})d\overline{y}}{\int_{y'}^{y^L} g_1(\overline{y})d\overline{y}} << \frac{\int_{y^L}^{y^L+d\overline{y}_{\phi}^M} \overline{P^{vat}} \cdot g_1(\overline{y})d\overline{y}}{\int_{y^L}^{y^L+d\overline{y}_{\phi}^M} g_1(\overline{y})d\overline{y}} \quad . \tag{34}$$

C Additional Institutional Background

LTU Eligibility Rule

The Spanish Tax Authority fixes a yearly revenue criteria to allocate firms under LTU monitoring. In particular, firms report to the tax agency their total revenue of year t in January of $t+1$ when their annual VAT summary must be submitted. Firms with revenue in year t that exceeds the €6.01 million threshold are monitored by the LTU since $t+1$, and the LTU is also in charge of reviewing all their tax returns from year t (e.g. CIT, VAT, wages withholding). Similarly, if revenue falls below the threshold in year s, the firm is removed from the LTU census in $s+1$ and tax returns from year s are not monitored by the LTU.

Groups

According to Spanish tax and business laws, there are two types of business groups: corporate groups and consolidated fiscal groups. The latter are more narrowly defined than the former, which require sharing the same activity and that the dominant firm owns at least 75% of the subsidiary's capital (see article 67 of the Royal Decree 4/2004 of the CIT for details). The LTU's revenue criterion refers to individual legal entities or consolidated fiscal groups (article 121 of the VAT Law), not to corporate groups. Therefore, a firm with annual revenue below €6.01 million that belongs to a large corporate group will not be included in the LTU census, but it would be included if it were part of a consolidated fiscal group. Note that consolidated fiscal groups are typically formed by the largest corporations, whereas corporate groups are more common. We do not have access to data on which firms belong to fiscal groups, but according to the Statistics published by the Spanish Tax Authority from 2004 to 2007, less than 1% of firms in the revenue range €1.5-€6 million are considered part of the LTU due to belonging to fiscal groups. Since this is a very small share, we do not expect that including these firms in the analysis will introduce a noticeable bias in our bunching estimates.

Exceptions to LTU Eligibility Rule

Exporting firms that claim a VAT refund are automatically included in the LTU census, regardless of their operating revenue. We do not have data on VAT claims related to exports that allows us to identify these firms accurately, so we cannot exclude these firms from the analysis nor can we use this set of firms as a comparison group.

Two regions in Spain, Navarra and País Vasco, have their own independent tax authorities. Firms with headquarters located in each of these regions are monitored by those independent tax authorities, unless they obtain more than 75% of their operating revenue from transactions in other Spanish regions, in which case they are monitored by the national LTU. Since we do not have information on the geographic destination of sales at firm level, we are unable to identify which large firms in these regions are within the LTU stricter tax monitoring. The distribution of reported revenue features modest, but statistically significant, bunching in the two regions. We choose to exclude them from the main analysis because of the uncertainty about how many firms are subject to the LTU and also because they represent a small proportion (7.2%) of the total number of firms with revenue between €3 and €9 million.

Corporate Income Tax Threshold

The standard rate in the corporate income tax was 35% of taxable profits in the period 1995-2007. A lower rate of 30% was applied to firms under a revenue threshold that was modified over time: from €1.5 million in 1999 up to €10 million in 2010 (full details provided in Table A.2). The cutoff for this tax break overlapped with the LTU threshold in 2004, but was different in the rest of the years. The lower rate was applied only to the first €90,121 of taxable profits (€120,202 since 2005) creating a notch for eligible firms with low taxable profits, and a kink for those with high profits.

External Audit and Abbreviated Returns Threshold

Firms are required by law to have their annual accounts audited by an external private firm if they fulfill two of the following criteria for two consecutive years: (i) annual revenue above €4.75 million; (ii) total assets above €2.4 million;[57] and (iii) more than 50 employees on average during the year. These criteria also determine whether a firm can use the abbreviated form of the corporate income tax return, rather than the standard (long) version. These requirements create compliance costs,[58] and the private audit information could complement tax enforcement. Private auditors have a legal responsibility to communicate tax misreporting to the authorities only in the (extreme) case of detecting

[57] The revenue limit was originally 790 million pesetas (€4.748 million), and the assets limit was 395 million pesetas (€2.374 million).

[58] The yearly fee charged by private audit firms is in the range €10,000 - €30,000 for firms with revenue close to €4.75 million, a small but non-negligible expenditure (0.2 to 0.6% of total revenue, but 4 to 12% of reported *profits* on average).

systematic fraud and criminal activities. However, they do not have to report neither their statement nor detected accounting inconsistencies to the tax authority.

Third-Party Information Reported in Tax Returns

The tax authority introduced in the 1980's a mandatory information form (*Modelo 347*) in which all firms, both below and above the LTU threshold, must provide detailed information on the monetary value of their transactions with all of their suppliers and clients. The information from these forms is processed electronically and regularly used by tax auditors to cross-check tax returns and detect discrepancies between trading partners. Note that we do not have access to any data from these forms.

D Data: Further Details

Main Variables Used in the Analysis

The main variables used in our empirical analysis are: (i) annual net operating revenue, which is used to determine whether firms are eligible to the LTU; (ii) material expenditures, i.e. the cost of all raw materials and services purchased by the firm in the production process; (iii) labor expenditures, which accounts for the total wage bill of a firm, excluding social security contributions; (iv) number of employees; (v) accounting profit, i.e. the gross profit reported in the CIT; (vi) actual tax liability in the CIT, and (vii) taxable profit, which we calculate by applying the CIT schedule for firms that report a positive CIT liability. Table A.5 reports summary statistics for all these variables.

Definition of Sectors of Activity

Table A.1 provides the sector definitions that we use in section 5.2 in terms of the 2009 version of the National Classification of Economic Activities (in Spanish, CNAE), which follows the Eurostat standard NACE Rev. 2. We use 2-digit CNAE codes to define sectors. The third column shows the number of firms in each sector for the 1995-2007 pooled CBB dataset, and the last column shows the percentage they represent overall.

Original CBB data and final dataset

We start from the original CBB data as provided by the *Banco de España* in September 2014. We include data for the years 1995 through 2007, both included. In order to construct the final dataset for our analysis, we take several steps. First, we drop observations from two regions where tax collection is independent of the federal tax authority and hence the LTU threshold does not apply (País Vasco and Navarra). Second, we choose a bin width of €60,101, which is one-hundredth of the revenue level of the LTU threshold. For symmetry, we keep 50 bins below and 50 bins above the threshold, so in total there are 100 bins. Hence, our final dataset has firms with reported revenue between €3.005 million and €9.015 million. Within this range, we define some of the ratios that we use in the section on input expenditures: materials and labor expenditures as % of revenue, average gross wages (defined as the total wage bill divided by the number of employees), and fixed assets as % of revenue. Finally, we drop the top and bottom 1% of observations from each of these variables, in order to avoid the presence of outliers in the data. There is some overlap in the extreme values, such that a firm with abnormally high materials is likely to have abnormally low labor expenditures. The final dataset contains 285,570 observations, and summary statistics are reported on Table A.5. The Stata do-files used to process the original data to arrive at the final dataset are available upon request.

Input-Output Tables

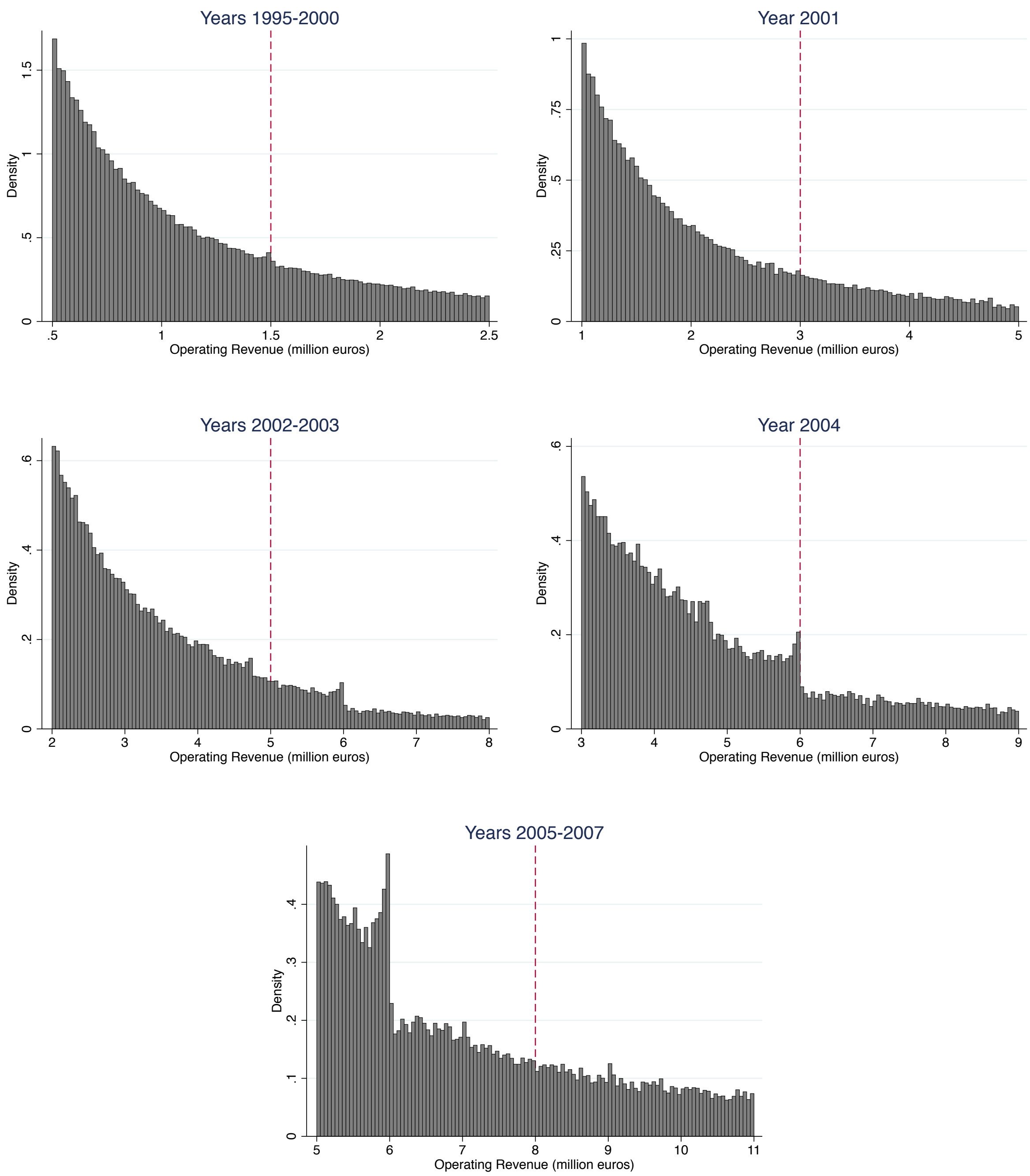
We use the input-output tables produced by the National Statistics Institute (*Instituto Nacional de Estadistica, INE*) for the year 2000. Sectors of activity are defined according to Spanish industry classification (TSIO), which does not match CNAE 2009 codes exactly but has substantial overlap. To calculate the share of sales made to final consumers by sector, we divide the column labelled *"Consumo final de los hogares, interno"* ("Households' final consumption, domestic") by the column *"Total empleos"* ("Total uses"). The original table used for the calculations can be downloaded from:

www.ine.es/daco/daco42/cne00/simetrica2000.xls

The table we provide together with our main dataset contains, additionally, the correspondence between our sector definitions (based on CNAE 2009 codes) and the sectors defined in the input-output tables.

Appendix Figures

Figure A.1: Behavioral (Non)response at the Corporate Income Tax Threshold



Notes: these graphs show the operating revenue distribution for different periods, around the threshold for the corporate income tax cut for small firms. Only firms with positive taxable profits are included (about 80% of the full sample), because the tax rate is irrelevant for them. There is no bunching at this threshold in any year except for 2004, the year in which this cutoff overlapped with the LTU threshold discussed in the main text. The results are essentially identical when using the full sample of firms.

Figure A.2: Distribution of Reported Revenue, by Year

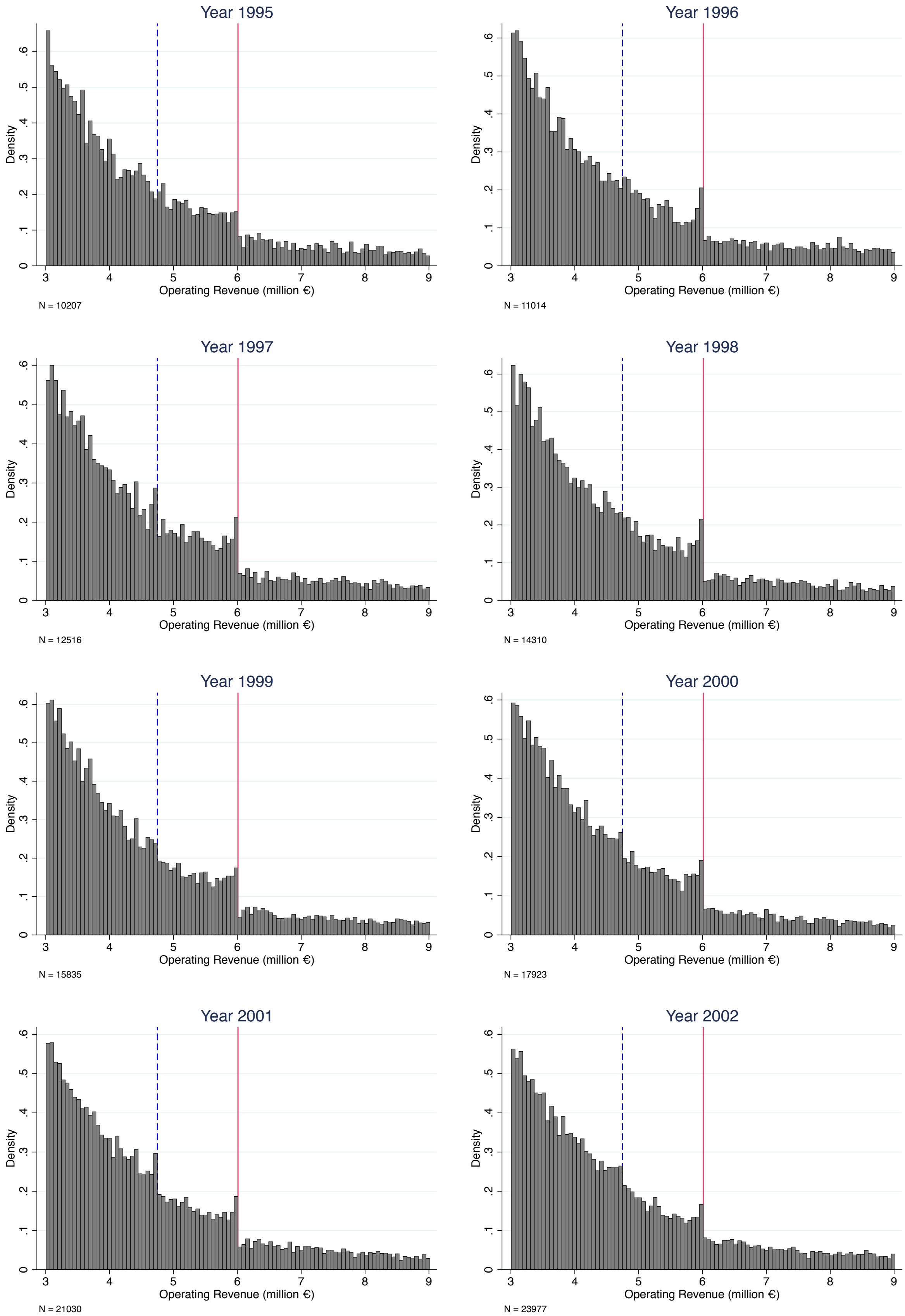
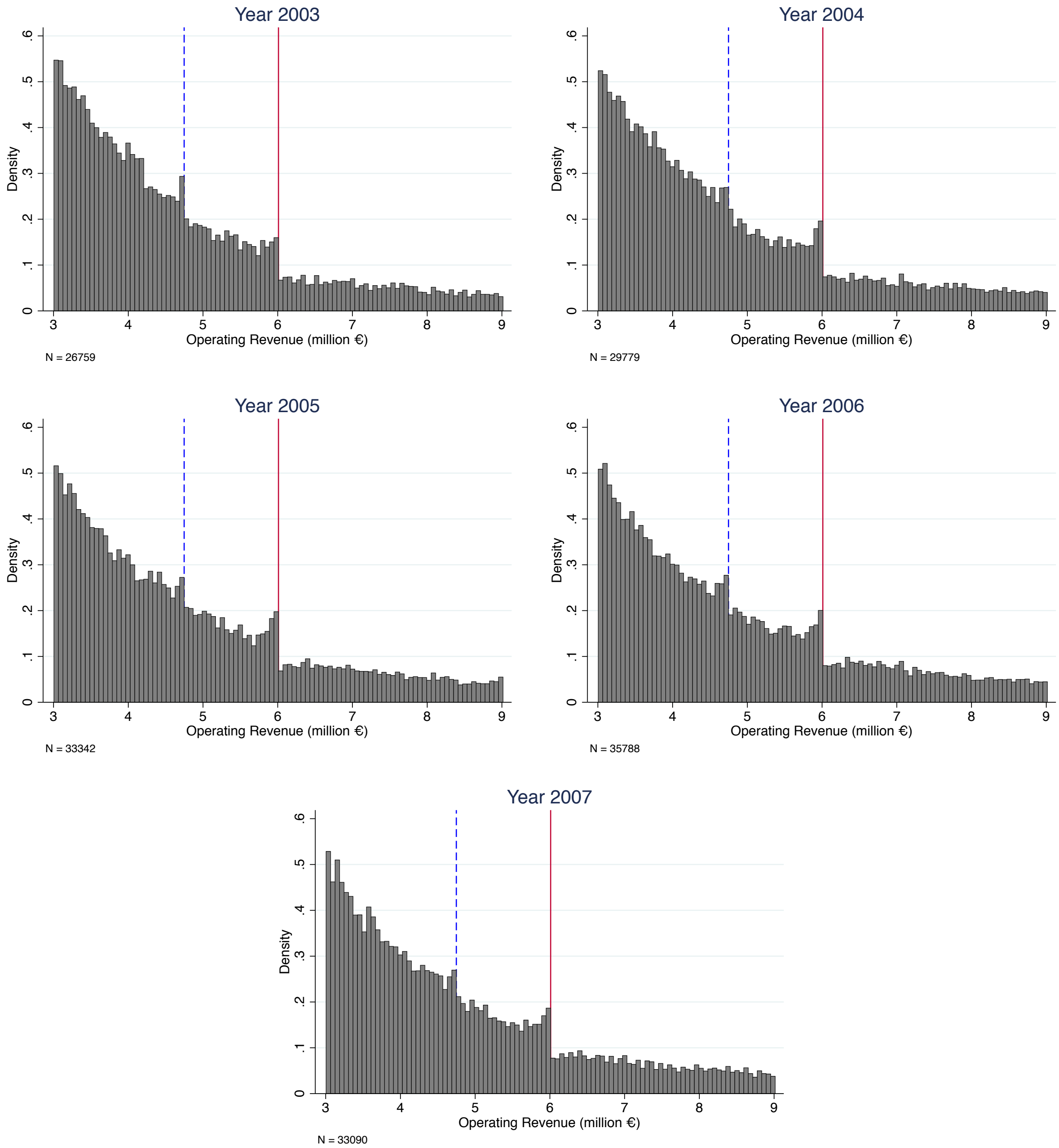


Figure A.2: Distribution of Reported Revenue, by Year (continued)



Notes: these graphs show the operating revenue distribution for each year in the period under study (1995-2007).

Figure A.3: Bunching Response by Sector

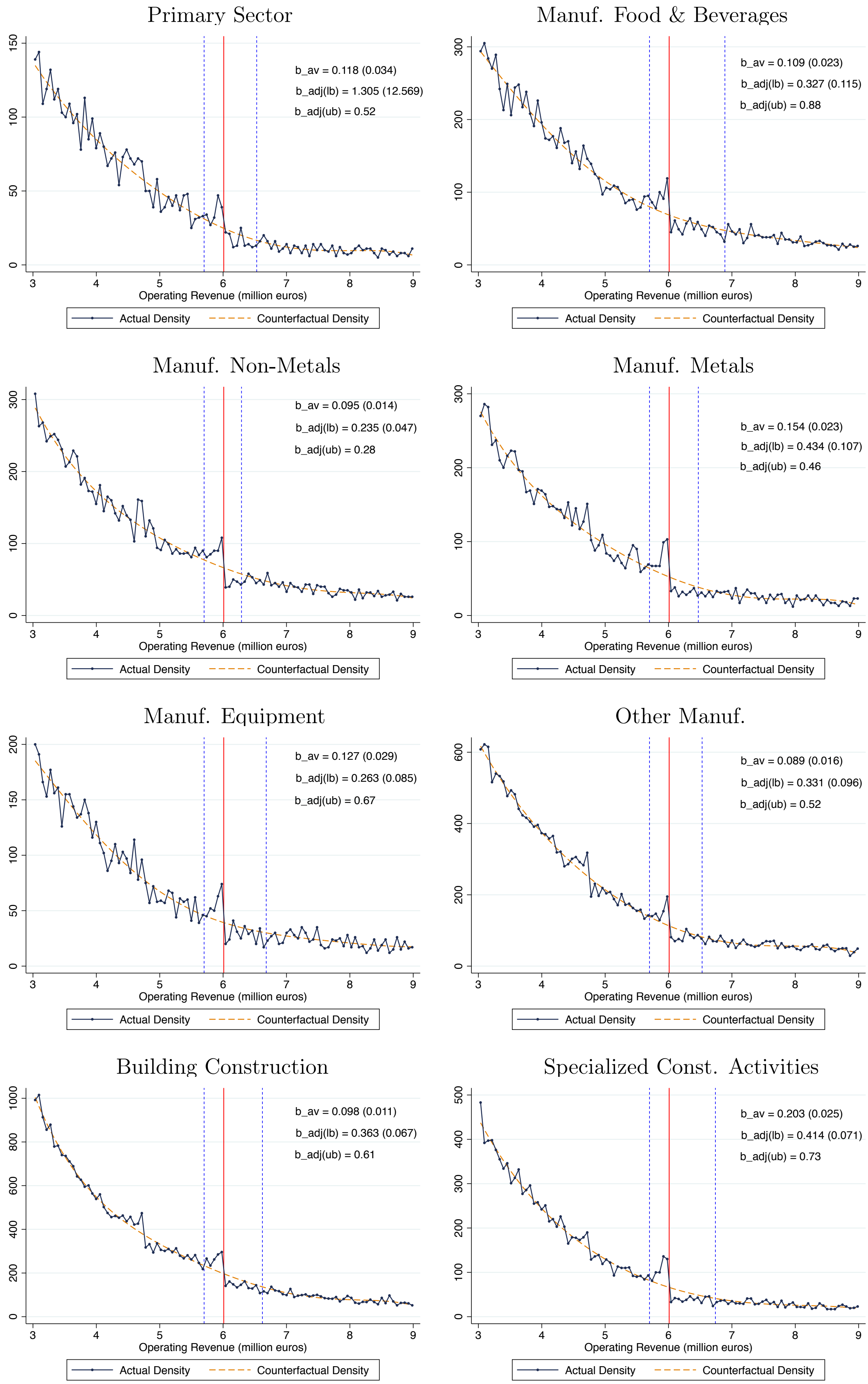
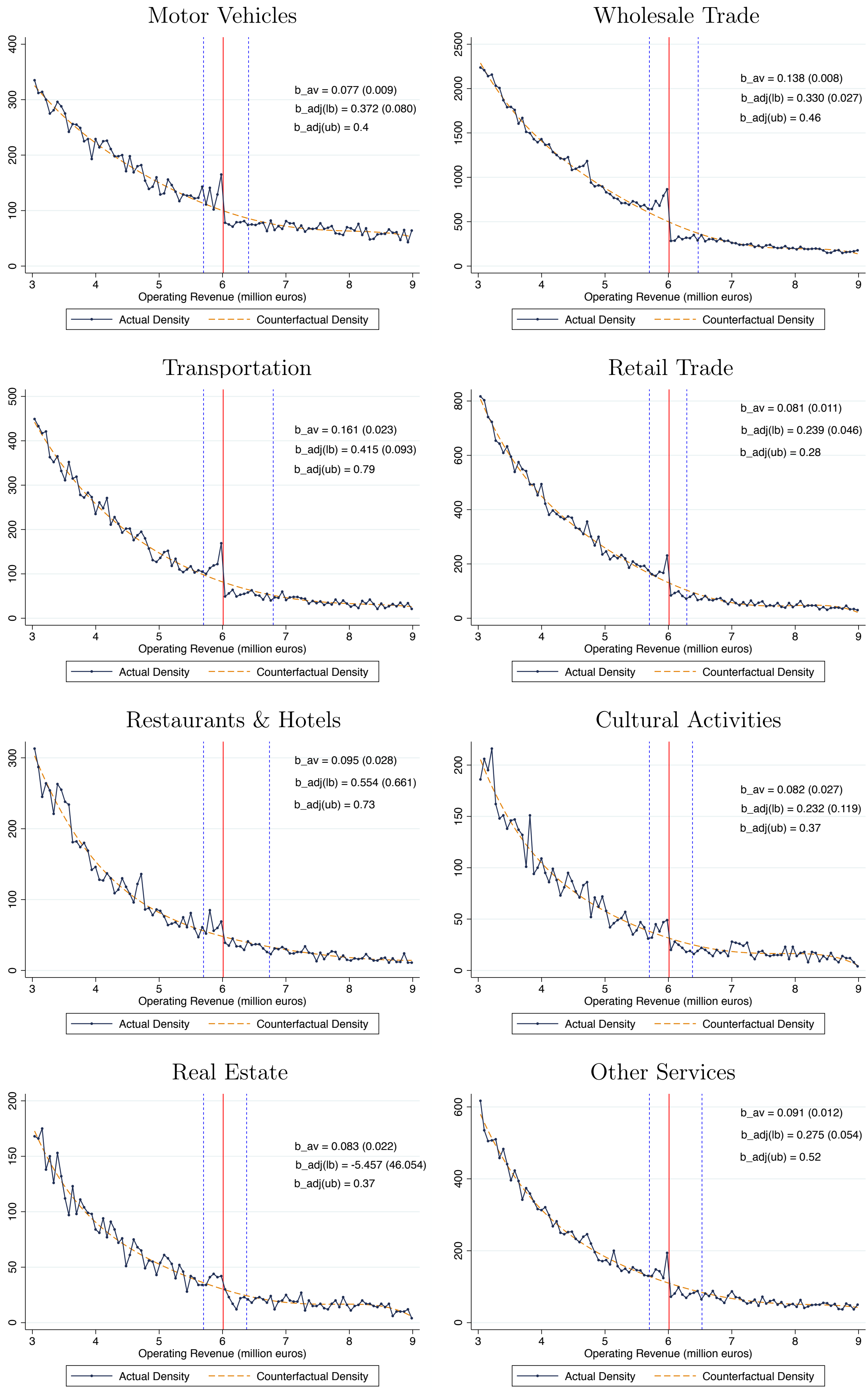


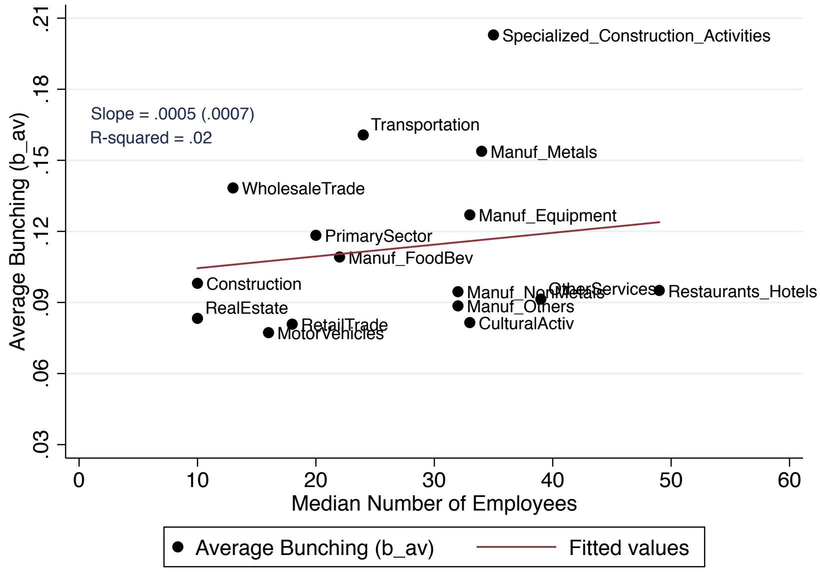
Figure A.3: Bunching Response by Sector (continued)



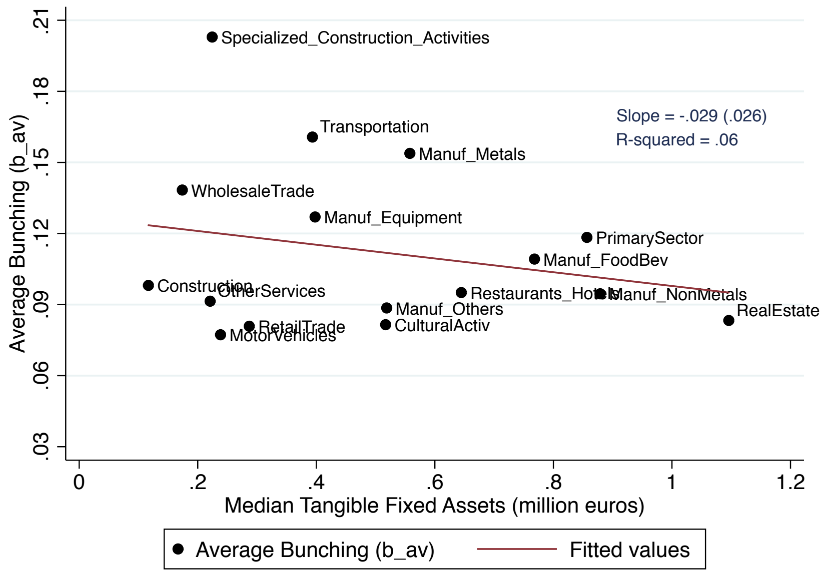
Notes: these graphs show the observed and counterfactual operating revenue distribution for each sector in the period 1995-2007. The dashed (red) vertical line indicates the LTU threshold. The estimation method is identical to that applied in Figure 2 and explained in the main text.

Figure A.4: Average Bunching by Firm Size Characteristics

(a) Median Number of Employees



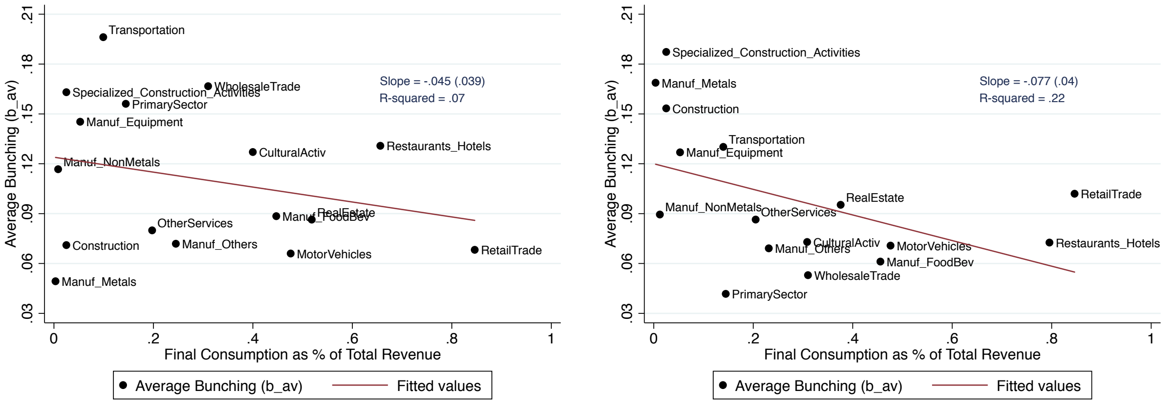
(b) Median Tangible Fixed Assets



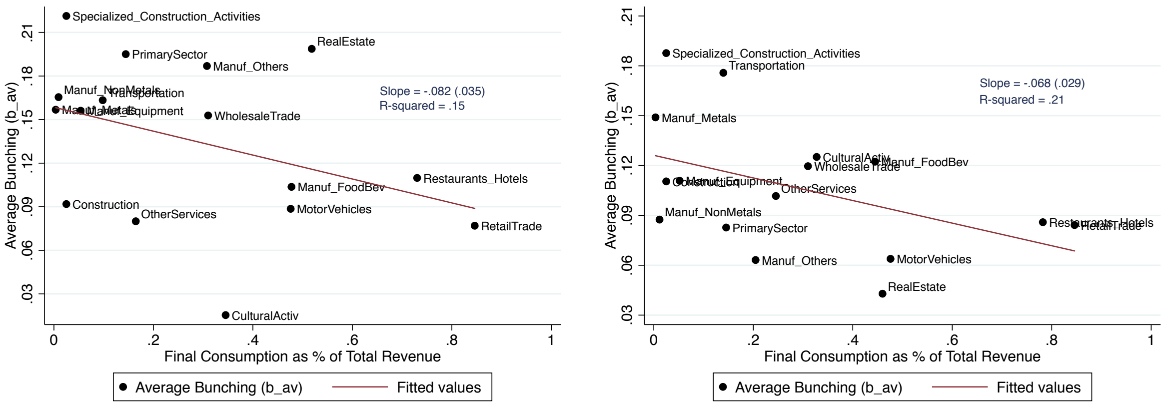
Notes: these graphs are robustness checks for the complementarity result depicted in Figure 3. The top panel shows average bunching against the median number of employees by sector. The bottom panel shows average bunching against median tangible fixed assets by sector. The slope of the relationship is close to zero and not statistically significant in either case, suggesting that firm size characteristics such as employment or tangible assets are not strongly related to bunching behavior at the sector level.

Figure A.5: Robustness of Complementarity Result

(a) Below (left) or Above (right) Median Number of Employees



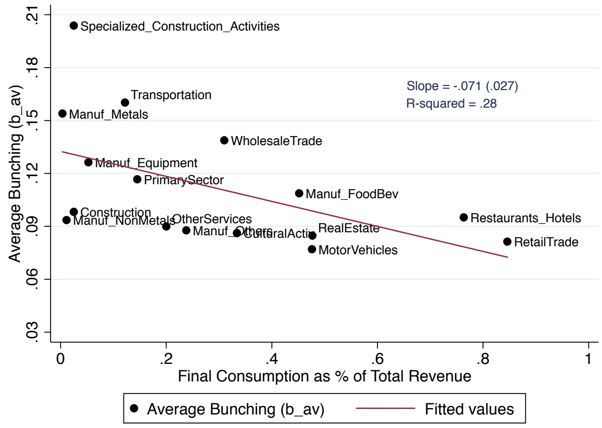
(b) Below (left) or Above (right) Median Tangible Fixed Assets



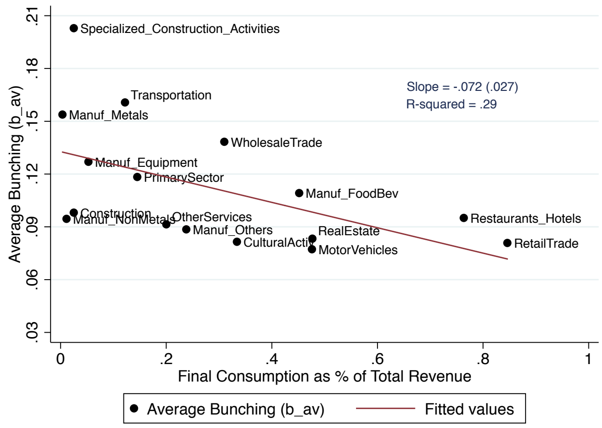
Notes: these graphs are robustness checks for the complementarity result depicted in Figure 3. The top panel shows the average bunching estimates by sector, using only firms below (left panel) and above (right panel) the median number of employees in the overall sample. Similarly, the bottom panel shows the average bunching estimates by sector, using only firms below (left panel) and above (right panel) the median of tangible fixed assets in the overall sample. The slope of the relationship is negative in all cases as in Figure 3, and it is statistically different from zero in all cases except for the firms below the median number of employees.

Figure A.6: Robustness of Complementarity Result to Bin Size

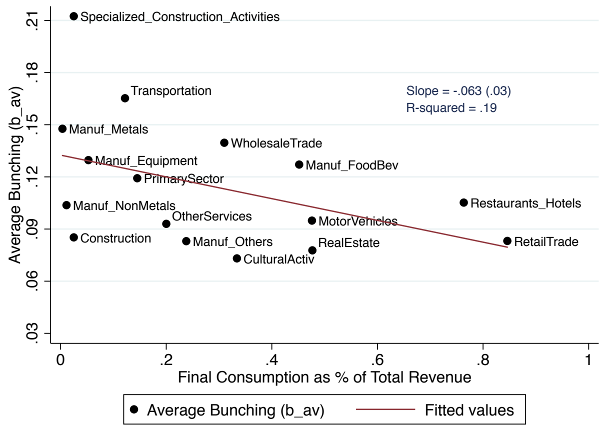
(a) Bin width = €30,050



(b) Bin width = €60,101 (as reported in Figure 3)

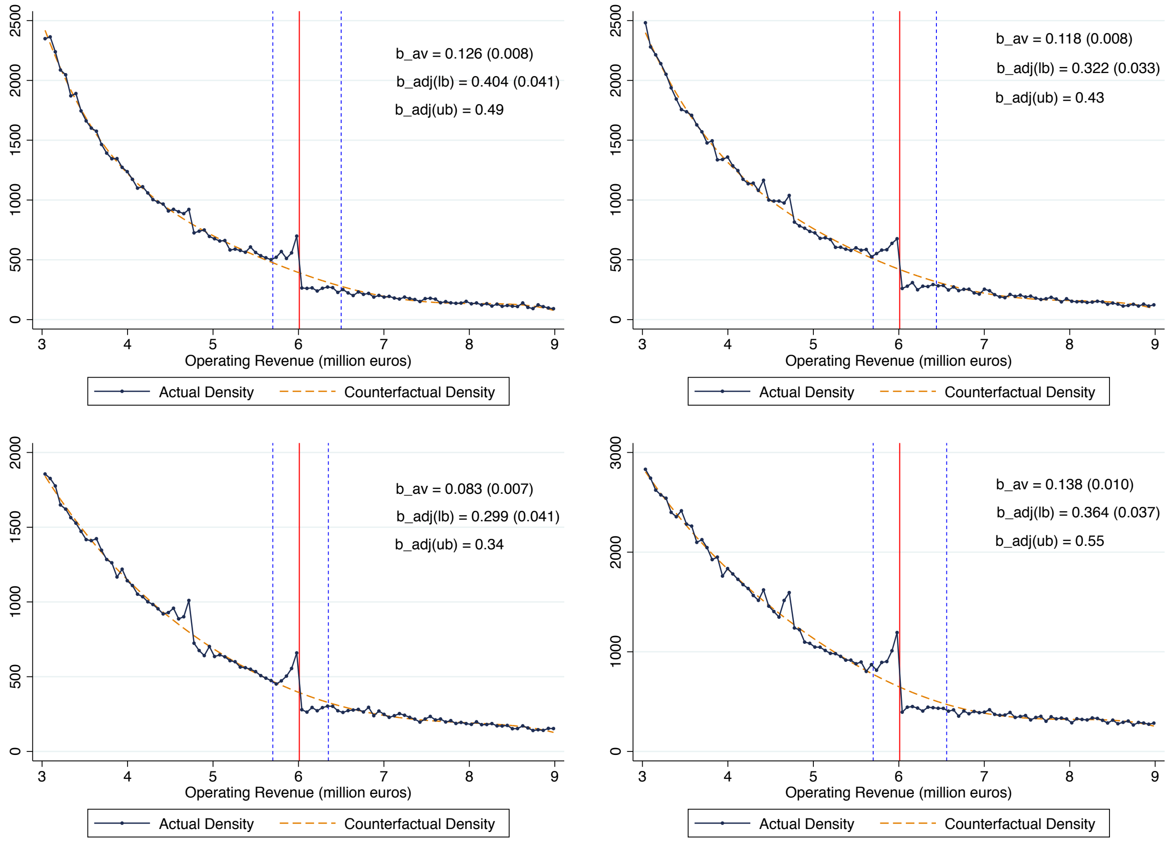


(c) Bin width = €120,202



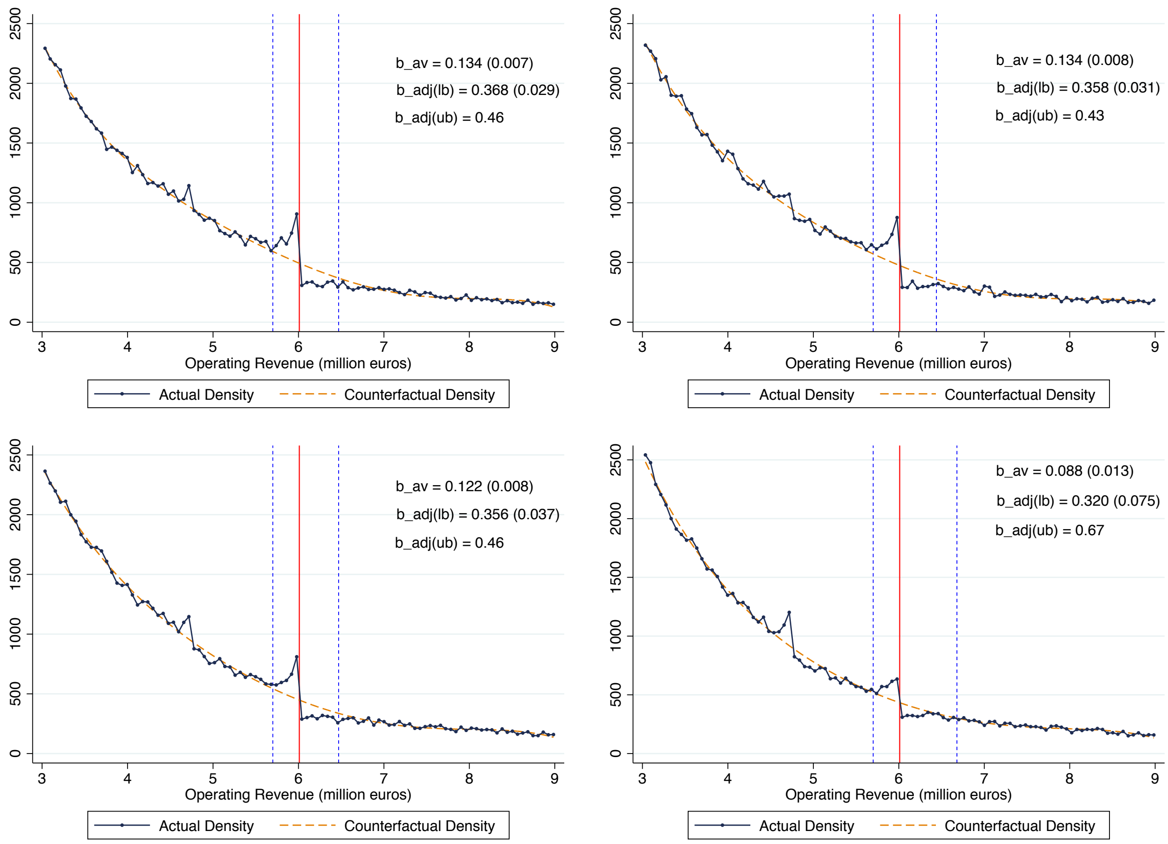
Notes: these graphs are robustness checks for the complementarity result depicted in Figure 3, showing how the results change using different bin sizes. The top panel shows the correlation between average bunching by sector and share of final sales using a bin width of €30,050 in the bunching estimation. The middle panel uses a bin width of €60,101 (exactly the same as Figure 3). The bottom panel uses a bin width of €120,202. The slope of the relationship is negative and significant in all cases, and the magnitude is very similar.

Figure A.7: Firm Size Distribution by Quartile of Employment



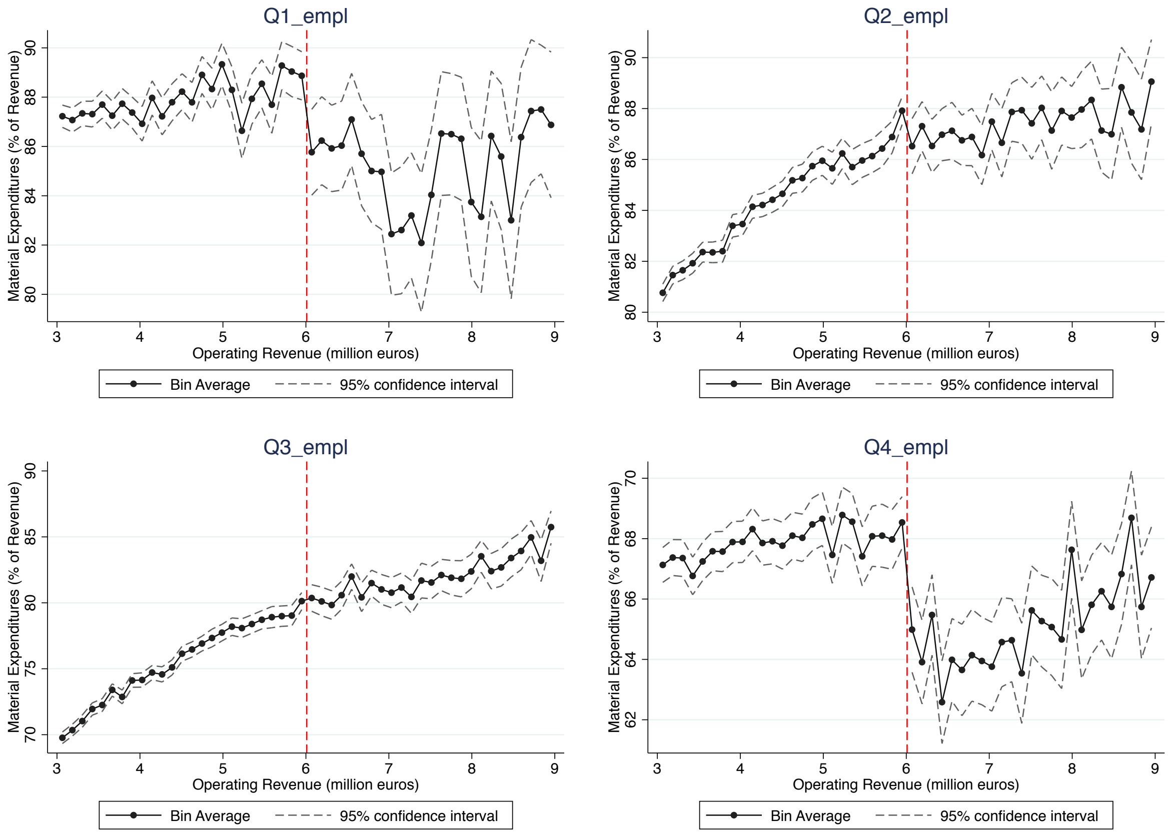
Notes: these graphs show the observed and counterfactual operating revenue distribution by quartiles of the employment distribution in the period 1995-2007. We divide all firms in the sample into four quartiles based on the employment distribution. Q1 = 0-9 employees; Q2 = 10-21 employees; Q3 = 22-38 employees; Q4 = 39 or more employees. The dashed (red) vertical line indicates the LTU threshold. The estimation method is identical to that applied in Figure 2 and explained in the main text. (Bin width= €60,101).

Figure A.8: Firm Size Distribution by Quartile of Tangible Fixed Assets



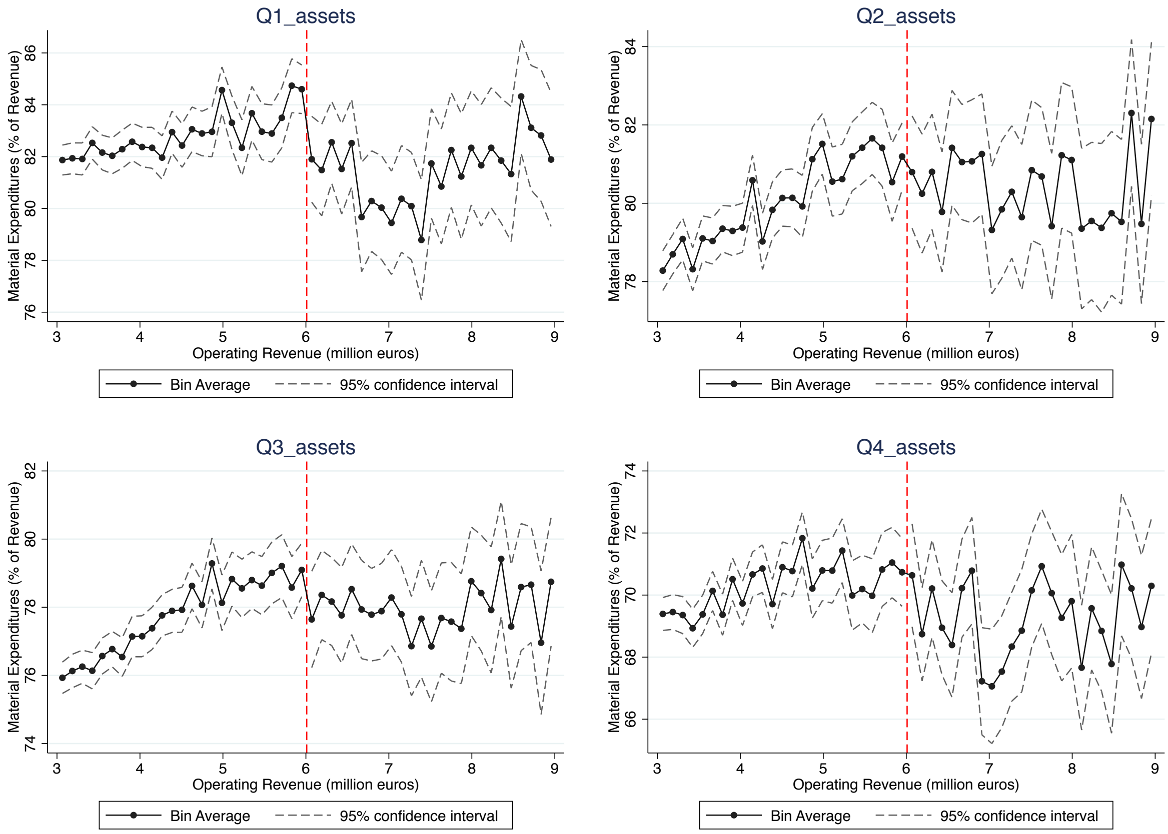
Notes: these graphs show the observed and counterfactual operating revenue distribution by quartiles of tangible fixed assets in the period 1995-2007. We divide all firms in the sample into four quartiles based on the distribution of tangible fixed assets. Q1 = (€0, €0.154) million in fixed assets; Q2 = (€0.145,€0.463); Q3 = (€0.463,€1.163); Q4 = (€1.163,∞). The dashed (red) vertical line indicates the LTU threshold. The estimation method is identical to that applied in Figure 2 and explained in the main text. (Bin width= €60,101).

Figure A.9: Reported Material Expenditures by Quartiles of Employment



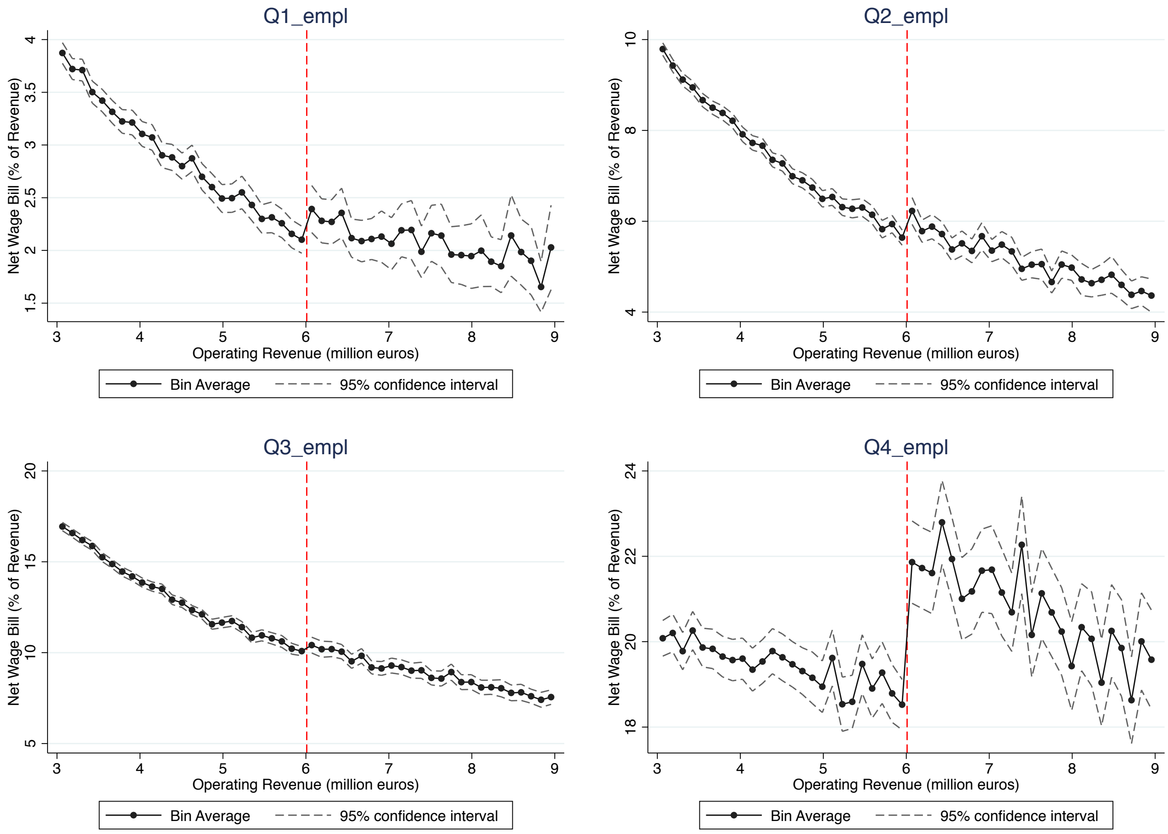
Notes: these graphs show the average ratio of material expenditures as a percentage of total revenue by sector of activity, the same outcome as Figure 4a in the main text. We divide all firms in the sample into four quartiles based on the employment distribution. Q1 = 0-9 employees; Q2 = 10-21 employees; Q3 = 22-38 employees; Q4 = 39 or more employees. The dashed (red) vertical line indicates the LTU threshold. The black dotted lines denote bin averages and the grey dashed lines show 95% confidence intervals for each bin average. To avoid the spurious effect of extreme values, we winsorize observations in the top and bottom 1% of the outcome variable, meaning that we set those values equal to the 1st and 99th percentile. We do this for each €1-million interval in the range $\overline{y} \in (3, 9)$ million. (Bin width= €120,202).

Figure A.10: Reported Material Expenditures by Quartiles of Tangible Fixed Assets



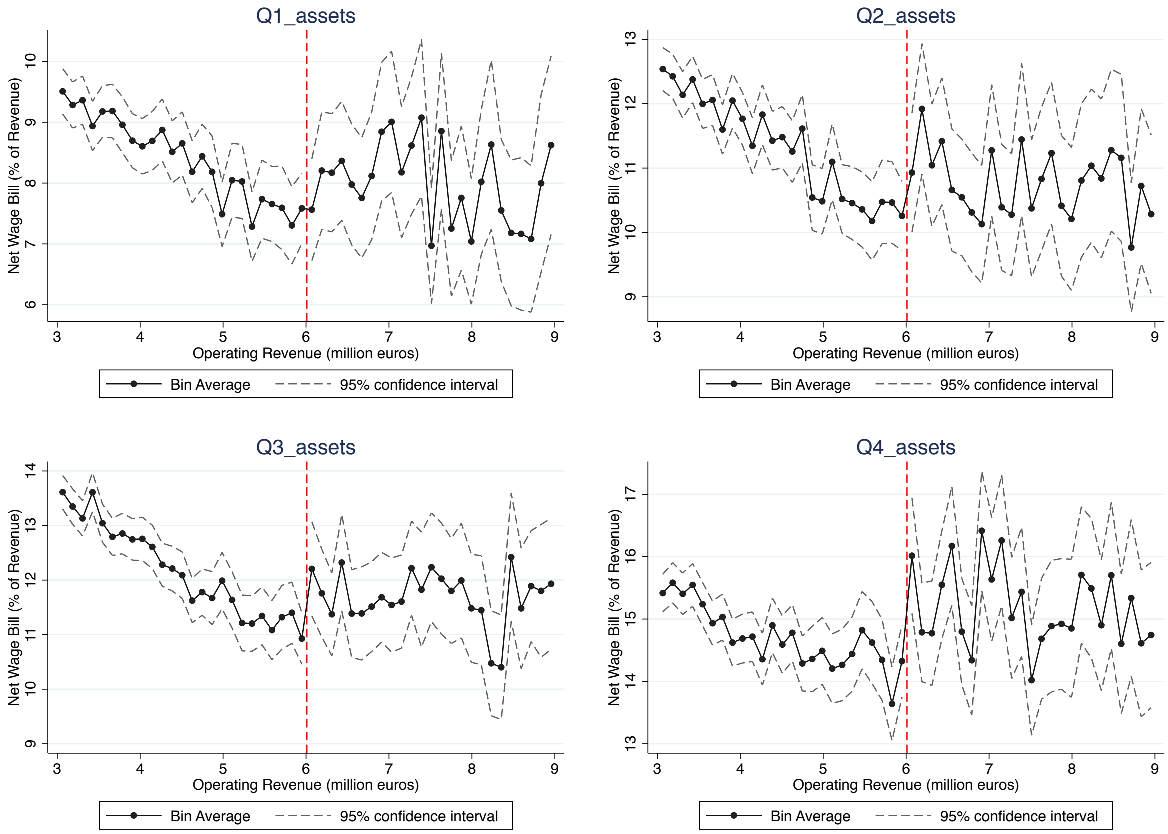
Notes: these graphs show the average ratio of material expenditures as a percentage of total revenue, the same outcome as Figure 4a in the main text. We divide all firms in the sample into four quartiles based on the distribution of total fixed assets. Q1 = (€0, €0.154) million in fixed assets; Q2 = (€0.145,€0.463); Q3 = (€0.463,€1.163); Q4 = (€1.163,∞). The dashed (red) vertical line indicates the LTU threshold. The black dotted lines denote bin averages and the grey dashed lines show 95% confidence intervals for each bin average. To avoid the spurious effect of extreme values, we winsorize observations in the top and bottom 1% of the outcome variable, meaning that we set those values equal to the 1st and 99th percentile. We do this for each €1-million interval in the range $\overline{y} \in (3, 9)$ million. (Bin width= €120,202).

Figure A.11: Reported Labor Expenditures by Quartiles of Employment



Notes: these graphs show the average ratio of the net wage bill as a percentage of total revenue. The net wage bill is defined as the total wage bill excluding employee-contributed payroll taxes (social security contributions), as in Figure 4b in the main text. We divide all firms in the sample into four quartiles based on the employment distribution. Q1 = 0-9 employees; Q2 = 10-21 employees; Q3 = 22-38 employees; Q4 = 39 or more employees. The dashed (red) vertical line indicates the LTU threshold. The black dotted lines denote bin averages and the gray dashed lines show 95% confidence intervals for each bin average. To avoid the spurious effect of extreme values, we winsorize observations in the top and bottom 1% of the outcome variable, meaning that we set those values equal to the 1st and 99th percentile. We do this for each €1-million interval in the range $\overline{y} \in (3, 9)$ million. (Bin width= €120,202).

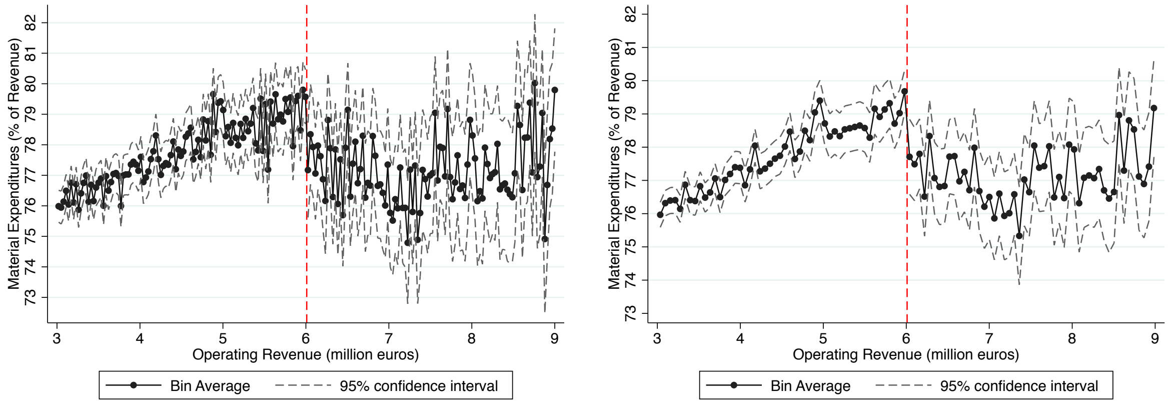
Figure A.12: Reported Labor Expenditures by Quartiles of Tangible Fixed Assets



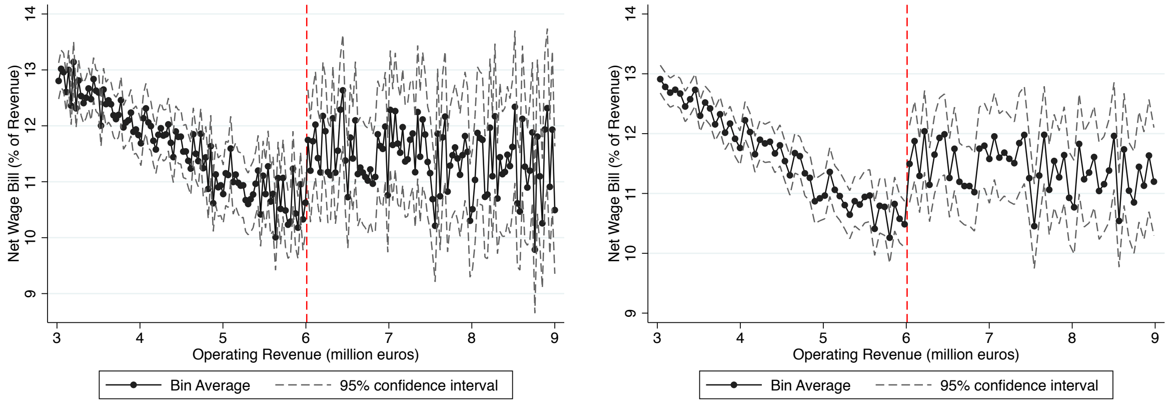
Notes: these graphs show the average ratio of the net wage bill as a percentage of total revenue by sector of activity. The net wage bill is defined as the total wage bill excluding employee-contributed payroll taxes (social security contributions), as in Figure 4b. We divide all firms in the sample into four quartiles based on the distribution of total fixed assets. Q1 = (€0, €0.154) million in fixed assets; Q2 = (€0.145,€0.463); Q3 = (€0.463,€1.163); Q4 = (€1.163,∞). The dashed (red) vertical line indicates the LTU threshold. The black dotted lines denote bin averages and the gray dashed lines show 95% confidence intervals for each bin average. To avoid the spurious effect of extreme values, we winsorize observations in the top and bottom 1% of the outcome variable, meaning that we set those values equal to the 1st and 99th percentile. We do this for each €1-million interval in the range $\overline{y} \in (3, 9)$ million. (Bin width= €120,202).

Figure A.13: Reported Input Expenditures: Different Bin Sizes

(a) Reported Material Expenditures

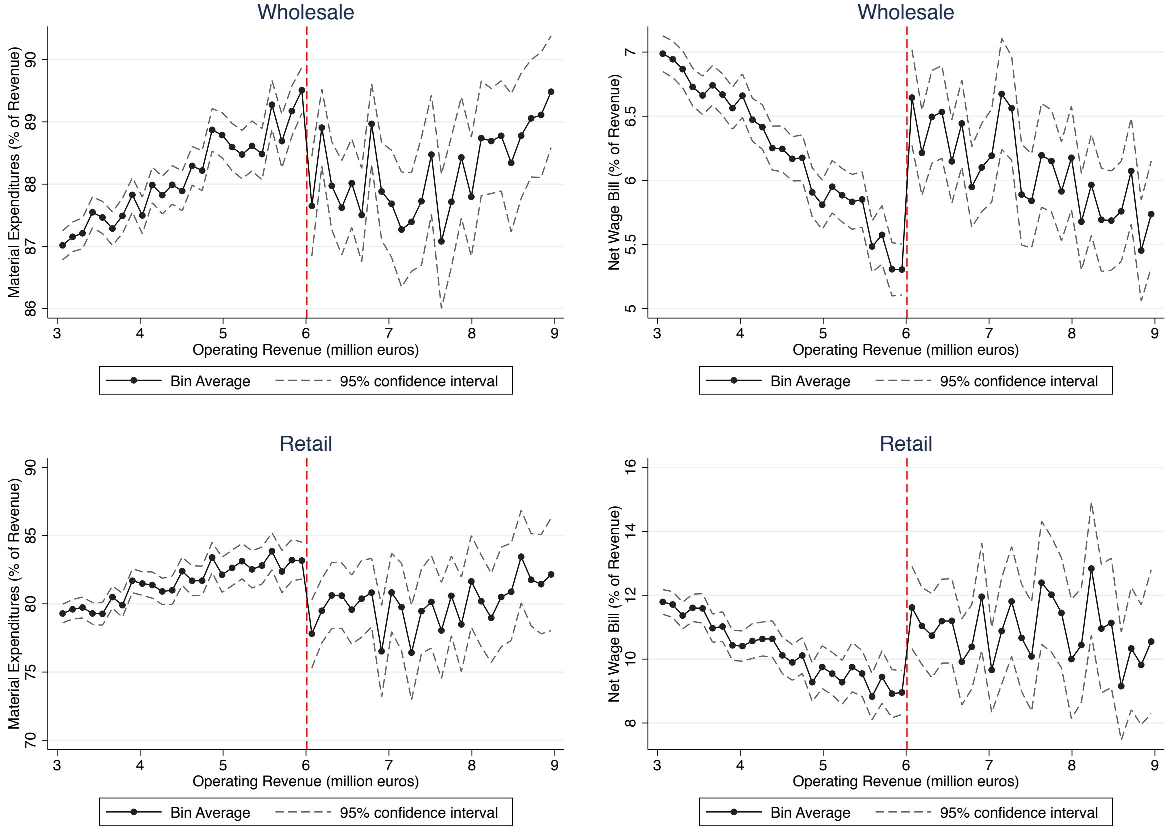


(b) Reported Labor Expenditures



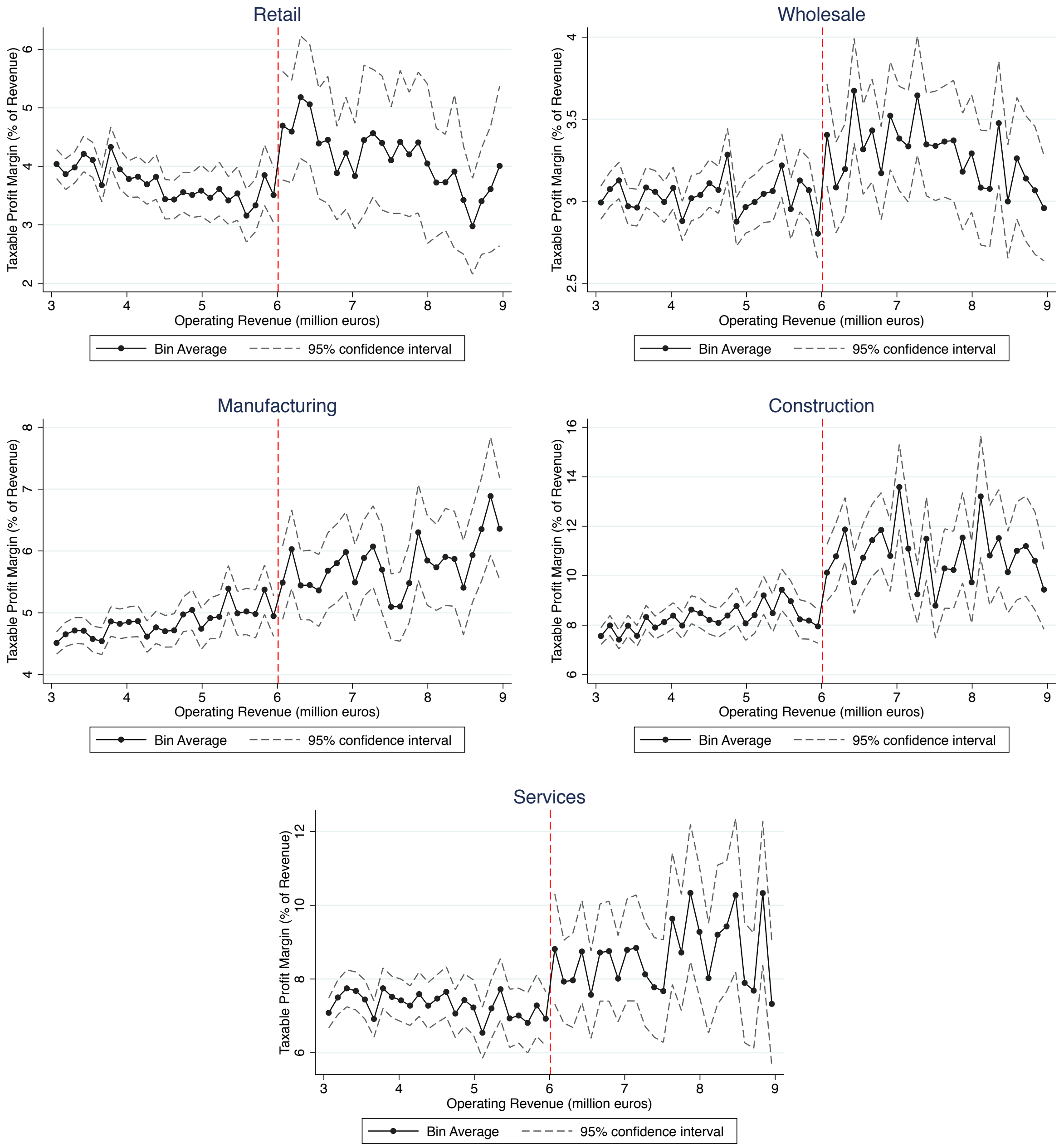
Notes: these graphs show the averages of reported inputs (materials and labor) using smaller bin sizes than in Figure 4. For each outcome, the figure on the left uses a bin width of €30,050, and the figure on the right uses a bin width of €60,101. The patterns are essentially the same as in the figure from the main text, where the bin size is €120,202.

Figure A.14: Reported Input Expenditures: Wholesale vs. Retail



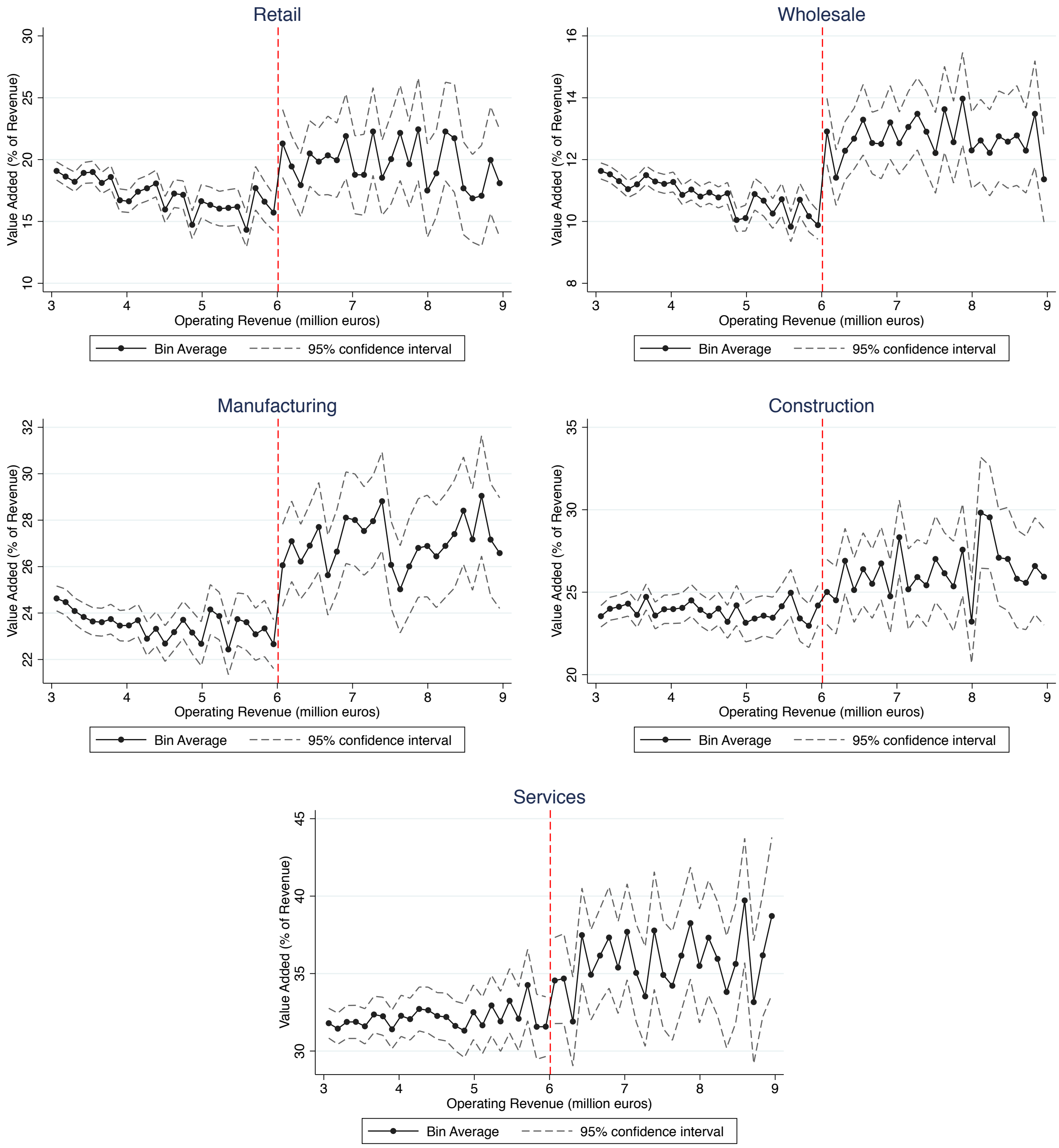
Notes: these graphs show average material and labor expenditures (the same outcomes as in Figure 4 in the main text) for two broad sectors, wholesale and retail. These sectors are chosen because they represent two polar cases in terms of the share of sales made to final consumers. The sector definitions are broader than in some of the previous exercises to ensure enough sample size. Wholesale includes wholesale trade and motor vehicles. Retail includes retail trade and restaurants and hotels. The dashed (red) vertical line indicates the LTU threshold. The black dotted lines denote bin averages and the gray dashed lines show 95% confidence intervals for each bin average. To avoid the spurious effect of extreme values, we winsorize observations in the top and bottom 1% of the outcome variable, meaning that we set those values equal to the 1st and 99th percentile. We do this for each €1-million interval in the range $\overline{y} \in (3, 9)$ million. (Bin width= €120,202).

Figure A.15: Reported Corporate Income Tax Base by Sector



Notes: these graphs show the average CIT tax base (taxable profit) as a percentage of total revenue by sector of activity. We distinguish 5 broad sectors of activity to ensure that there is enough statistical power to compare the behavior of firms below and above the LTU threshold, indicated by the the dashed (red) vertical line. The black dotted lines denote bin averages and the gray dashed lines show 95% confidence intervals for each bin average. To avoid the spurious effect of extreme values, we winsorize observations in the top and bottom 1% of the outcome variable, meaning that we set those values equal to the 1st and 99th percentile. We do this for each €1-million interval in the range $\overline{y} \in (3, 9)$ million. The "CIT tax base" is estimated as explained in the note to Figure 5. (Bin width= €120,202).

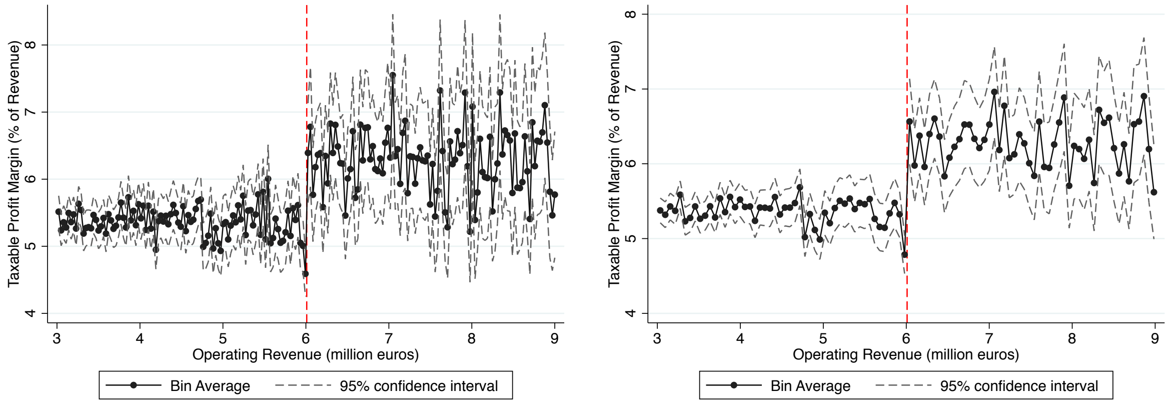
Figure A.16: Reported Value Added Tax Base by Sector



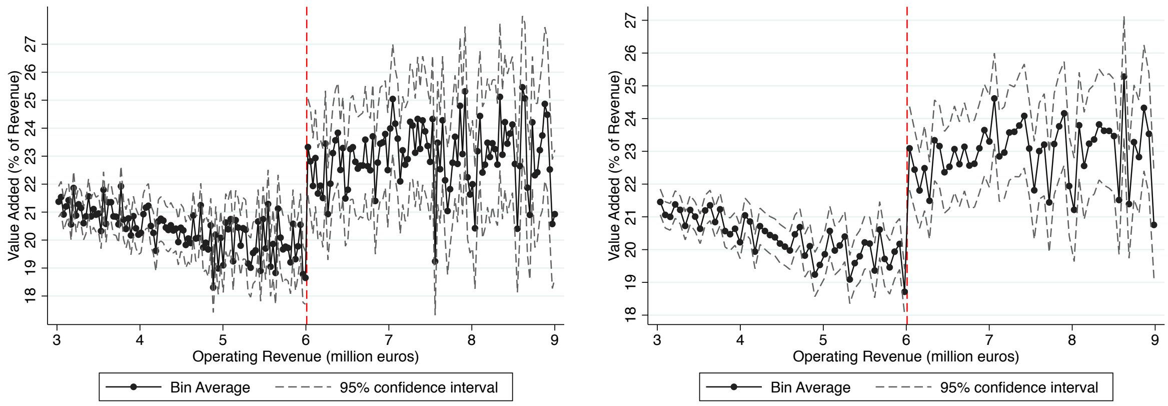
Notes: these graphs show the average value-added tax (VAT) base as a percentage of total revenue by sector of activity. We distinguish 5 broad sectors of activity to ensure that there is enough statistical power to compare the behavior of firms below and above the LTU threshold, indicated by the the dashed (red) vertical line. The black dotted lines denote bin averages and the gray dashed lines show 95% confidence intervals for each bin average. To avoid the spurious effect of extreme values, we winsorize observations in the top and bottom 1% of the outcome variable, meaning that we set those values equal to the first and 99th percentile. We do this for each €1-million interval in the range $\overline{y} \in (3, 9)$ million. (Bin width= €120,202).

Figure A.17: Reported Tax Bases: Different Bin Sizes

(a) Reported Taxable Profits (as % of Total Revenue)



(b) Reported Value Added (as % of Total Revenue)



Notes: these graphs show the averages of reported tax bases (taxable profits and value added) using smaller bin sizes than in Figure 5 in the main text. For each outcome, the figure on the left uses a bin width of €30,050, and the figure on the right uses a bin width of €60,101. The patterns are essentially the same as in Figure 5, where the bin size is €120,202. The dashed (red) vertical line indicates the LTU threshold. The black dotted lines denote bin averages and the gray dashed lines show 95% confidence intervals for each bin average. To avoid the spurious effect of extreme values, we winsorize observations in the top and bottom 1% of the outcome variable, meaning that we set those values equal to the first and 99th percentile. We do this for each €1-million interval in the range $\overline{y} \in (3, 9)$ million.

Appendix Tables

Table A.1: Number of Observations by 2-digit CNAE Sector codes

Sector	CNAE-2009 Sector Codes	# of Observations	
PrimarySector	01-09,19 - Agriculture, forestry, fishing, and mining	3,738	1.63%
Manuf_FoodBev	10,11,12 - Manufacture of food, beverages and tobacco	9,257	4.03%
Manuf_NonMetals	22,23 - Manufacture of plastics and non-metallic minerals	8,583	3,74%
Manuf_Metals	24,25 - Metal products, machinery	7,358	3.20%
Manuf_Equipment	26-28,33 - Manufacture of computers, electronics, equipment	5,506	2.40%
Manuf_Others	13-17,20,21,29-32 - Textiles, clothing, wood, paper, chemicals	17,367	7.56%
Const_Buildings	41 - Construction of buildings	25,888	11.27%
Const_SpecializedAct	43 - Specialized construction activities	10,327	4.50%
MotorVehicles	45 - Wholesale trade and repair of motor vehicles	12,134	5.28%
WholesaleTrade	46 - Wholesale trade (except motor vehicles)	66,406	28.92%
RetailTrade	47 - Retail trade	11,715	5.10%
RestHotels	55,56,79 - Hotels, restaurants and travel agencies	19,977	8.70%
Transportation	49-52 - Transportation by land, water, air, support activities	7,101	3.09%
CulturalActiv	18,58-60,90,93 - Publishing, movies, radio & TV, sports	4,934	2.15%
RealEstate	68,77- Real estate, rental and leasing	4,334	1.89%
OtherServices	53,61-64,69-75,78,80-82,85-88,92,95-96 - Other services	15,004	6.53%

Source: CBB dataset described in the main text for the number of observations. For the sector classifications, see http://www.ine.es/daco/daco42/clasificaciones/cnae09/estructuraen.pdf.

Table A.2: Revenue Threshold: Corporate Income Tax Benefit for Small Firms

<table>
<tr><th>Year</th><th>Threshold</th><th>Standard tax rate</th><th>Special tax rate</th><th>Applicable range</th></tr>
<tr><td>1999</td><td rowspan="2">€1.5 million</td><td rowspan="8">35%</td><td rowspan="8">30%</td><td rowspan="6">Up to €90,151 of taxable profits</td></tr>
<tr><td>2000</td></tr>
<tr><td>2001</td><td>€3 million</td></tr>
<tr><td>2002</td><td rowspan="2">€5 million</td></tr>
<tr><td>2003</td></tr>
<tr><td>2004</td><td>€6 million</td></tr>
<tr><td>2005</td><td rowspan="3">€8 million</td><td rowspan="3">Up to €120,202 of taxable profits</td></tr>
<tr><td>2006</td></tr>
<tr><td>2007</td><td>32.5%</td><td>27.5%</td></tr>
</table>

Source: the applicable laws are: Law 43/1995 (Article 122), Law 6/2000 (Article 122), Law 24/2001 (Article 122), Law 4/2004 (Article 108), Law 2/2004 of the *Presupuestos Generales del Estado* (Annual Government Budget Law, Article 108).

Table A.3: Overview of the Spanish Tax System

	Top tax rate	*Share of tax revenue*
Social Security Contributions (PRT)	38%	33%
Personal Income Tax (PIT)	48% (46%)	22%
Value-Added Tax (VAT)	16%	19%
Corporate Income Tax (CIT)	35% (30%)	13%
Other indirect taxes and fees	-	13%
Federal Tax Revenue / GDP	30-37%	

Sources: Instituto de Estudios Fiscales (2011). The top marginal rate of the individual income tax was reduced to 46% 2005. The top marginal rate of the corporate income tax was reduced to 32.5% in 2006 and 30% in 2007. The data on tax revenues reflects averages for the period 1999-2007 and includes regional-level revenues in all calculations.

Table A.4: CBB Dataset Compared to Official CIT Statistics

		All Firms		€3-€10 million	
1995	Official Statistics	564,146		20,686	
	CBB Database	435,482	77.9%	12,592	60.9%
1996	Official Statistics	607,186		22,216	
	CBB Database	483,028	80.4%	13,924	62.7%
1997	Official Statistics	651,510		23,892	
	CBB Database	530,590	82.2%	16,216	67.9%
1998	Official Statistics	700,169		25,659	
	CBB Database	591,974	85.3%	18,453	71.9%
1999	Official Statistics	743,660		26,199	
	CBB Database	604,744	81.3%	20,083	76.7%
2000	Official Statistics	823,659		31,294	
	CBB Database	635,627	77.2%	22,468	71.8%
2001	Official Statistics	872,713		34,391	
	CBB Database	726,119	83.2%	25,561	74.6%
2002	Official Statistics	942,148		37,157	
	CBB Database	813,516	86.3%	29,003	78.1%
2003	Official Statistics	971,756		39,786	
	CBB Database	879,042	90.5%	32,191	80.9%
2004	Official Statistics	1,042,725		43,062	
	CBB Database	953,153	91.4%	35,846	83.2%
2005	Official Statistics	1,121,879		46,977	
	CBB Database	1,024,183	91.3%	40,422	86.0%
2006	Official Statistics	1,267,542		52,396	
	CBB Database	1,054,238	83.2%	43,325	82.7%
2007	Official Statistics	1,330,911		55,843	
	CBB Database	1,068,001	80.2%	39,728	71.1%

Notes: The percentages indicate the proportion of firms with a legal status of *Sociedad Anónima* (SA, equivalent to Corporation) or *Sociedad Limitada* (SL, equivalent to Limited Liability Company) in the CBB dataset compared to the number of firms with the same legal status that submitted a corporate income tax return that year. Official statistics have been compiled by the fiscal division of *Banco de España* based on several issues of *"Memoria de Administración Tributaria"*, an annual report published by the Spanish tax agency (AEAT, 1995-2007). The CBB dataset is described in detail in section 4.

Table A.5: Summary Statistics

	Mean	SD	Median	Min.	Max.	Obs.
Operating Revenue (million €)	4.669	1.447	4.253	3.005	9.015	285,580
Material Expenditures (million €)	3.630	1.449	3.347	0.000	28.698	279,878
Net Wage Bill (million €)	0.520	0.530	0.369	0.000	11.017	260,884
Taxable Profits (million €)	0.245	0.356	0.116	0.002	5.295	237,180
CIT Liability (million €)	0.068	0.116	0.027	-0.644	1.826	279,879
Value Added (million €)	0.959	1.037	0.692	-6.325	33.579	280,371
Tangible Fixed Assets (million €)	1.041	1.979	0.455	0.000	138.412	282,477
Number of Employees (FTE)	27.8	28.1	20	0	429	247,884
Material Expenditures (% of Revenue)	77.7%	17.8%	82.0%	0.0%	358.7%	279,878
Net Wage Bill (% of Revenue)	11.2%	10.4%	8.3%	0.0%	122.4%	260,885
Taxable Profit Margin (% of Revenue)	5.17%	6.8%	2.63%	0.0%	86.6%	237,184
Value Added (% of Revenue)	20.4%	19.2%	15.9%	-70.2%	419.7%	280,374

Notes: this table shows summary statistics for firms in the final dataset used for analysis, which is restricted to firms with reported revenue $y \in$(€3.01,€9.01) million. The top and bottom 1% of the variables "Materials as % of revenue", "Labor as % of revenue", "Fixed assets as % of revenue" and "Average gross wage" were dropped from the initial dataset to prevent outliers (and potentially incorrect data entries) from biasing the empirical estimations. The number of observations is different for each variable due to missing values, an issue especially relevant for the number of employees variable, which is not reported by about 20% of the firms.
Source: annual data from the Banco de España's CBB dataset for Spanish firms in the period 1995-2007, built using administrative data from *Registro Mercantil*. More details about the dataset are given in online appendix D.

Table A.6: Sensitivity Analysis, Pooled 1995-2007 data

Polynomial degree q	Excluded Interval y_{lb}	y_{ub}	Bunching Estimators b_{av}	b_{adj}	Obs.
4	5.30	6.68	0.151	0.439	285,570
			(0.015)***	(0.079)***	
5	5.30	6.68	0.140	0.411	285,570
			(0.013)***	(0.072)***	
4	5.40	6.68	0.149	0.433	285,570
			(0.012)***	(0.068)***	
5	5.40	6.71	0.147	0.431	285,570
			(0.011)***	(0.064)***	
4	5.50	6.59	0.140	0.408	285,570
			(0.009)***	(0.050)***	
5	5.50	6.62	0.135	0.394	285,570
			(0.008)***	(0.048)***	
4	5.60	6.53	0.131	0.381	285,570
			(0.008)***	(0.041)***	
5	5.60	6.59	0.129	0.375	285,570
			(0.007)***	(0.040)***	
4	5.70	6.47	0.120	0.350	285,570
			(0.006)***	(0.035)***	
5	**5.70**	**6.53**	**0.121**	**0.382**	285,570
			(0.007)***	**(0.036)*****	
4	5.80	6.38	0.106	0.301	285,570
			(0.004)***	(0.023)***	
5	5.80	6.41	0.108	0.312	285,570
			(0.004)***	(0.020)***	

Notes: this table shows the sensitivity of the bunching estimators to different assumptions on the excluded region used to estimate the counterfactual and the order of the polynomial. In all rows, we use the pooled 1995-2007 sample including all firms with reported revenue $y \in$(€3.01,€9.01). We pick different values of q, as shown in the first column, and y_{lb}, as shown in the second column. We obtain the corresponding values for y_{ub} and the point estimates for the bunching estimators b_{av} and b_{adj} using the methods described in the main text. The results are very similar for all the reasonable choices of the lower bound (y_{lb}), and for polynomials of degree 4 and 5. We highlight the results for $y_{lb} = 5.70$ and $q = 5$, which are the values chosen to produce the main estimation results. Significance levels: *** = 1%, ** = 5%, and * = 10%.

Table A.7: Sensitivity of Adjusted Bunching Estimator by Year

	Upper Bound of Estimation Interval		
Year	$Y^D = 6.07$	$Y^D = 6.13$	$Y^D = 6.19$
Pooled data			
1995-2007	0.353	0.382	0.424
	(0.032)	(0.036)	(0.043)
Annual data			
1995	0.408	0.257	0.331
	(0.170)	(0.079)	(0.121)
1996	0.303	0.395	0.379
	(0.069)	(0.111)	(0.105)
1997	0.538	0.515	0.634
	(0.125)	(0.118)	(0.173)
1998	0.331	0.351	0.368
	(0.069)	(0.077)	(0.084)
1999	0.326	0.416	0.508
	(0.065)	(0.097)	(0.139)
2000	0.601	0.660	0.698
	(0.138)	(0.166)	(0.188)
2001	0.313	0.350	0.437
	(0.060)	(0.073)	(0.105)
2002	0.429	0.410	0.401
	(0.116)	(0.109)	(0.107)
2003	0.302	0.348	0.381
	(0.064)	(0.080)	(0.095)
2004	0.497	0.550	0.566
	(0.109)	(0.132)	(0.142)
2005	0.268	0.321	0.356
	(0.034)	(0.045)	(0.054)
2006	0.292	0.298	0.317
	(0.042)	(0.044)	(0.048)
2007	0.301	0.305	0.351
	(0.051)	(0.053)	(0.067)

Notes: this table shows the sensitivity of the adjusted bunching estimator b^{lb}_{adj} to different values of the upper bound of the interval on which this parameter is estimated, y^D. The main estimates reported in Table 1 are reported in the central column here. When $y^D = 6.07$, the proportion of non-bunchers α is estimated as the ratio of the counterfactual frequency to the actual frequency in the first bin to the right of the LTU threshold. When $y^D = 6.13$ and $y^D = 6.20$, we use the first two and three bins, respectively.

Table A.8: Sensitivity of Adjusted Bunching Estimator by Sector

Sector	Upper Bound of Estimation Interval $Y^D = 6.07$	$Y^D = 6.13$	$Y^D = 6.19$
Primary Sector	1.305	1.301	0.571
	(12.569)	(21.917)	(0.595)
Manuf. Food and Beverages	0.327	0.533	0.510
	(0.115)	(0.320)	(0.299)
Manuf. Non-Metals	0.235	0.245	0.294
	(0.047)	(0.050)	(0.067)
Manuf. Metals	0.434	0.528	0.450
	(0.107)	(0.152)	(0.117)
Manuf. Equipment	0.263	0.300	0.513
	(0.085)	(0.107)	(0.374)
Manuf. Others	0.331	0.291	0.306
	(0.096)	(0.079)	(0.087)
Construction of Buildings	0.363	0.478	0.499
	(0.067)	(0.105)	(0.114)
Specialized Constr. Activ.	0.414	0.496	0.529
	(0.071)	(0.097)	(0.109)
Motor Vehicles	0.372	0.361	0.345
	(0.080)	(0.077)	(0.072)
Wholesale (exc. Motor V.)	0.330	0.342	0.384
	(0.027)	(0.029)	(0.035)
Transportation	0.415	0.485	0.595
	(0.093)	(0.124)	(0.190)
Retail Trade	0.239	0.283	0.338
	(0.046)	(0.061)	(0.082)
Restaurants and Hotels	0.554	0.470	0.719
	(0.661)	(0.389)	(4.679)
Cultural Activities	0.232	0.393	0.449
	(0.119)	(0.439)	(2.375)
Real Estate	-5.457	0.952	0.456
	(46.054)	(5.939)	(0.729)
Other Services	0.275	0.326	0.455
	(0.054)	(0.071)	(0.125)

Notes: this table shows the sensitivity of the adjusted bunching estimator b_{adj}^{lb} to different values of the upper bound of the interval on which this parameter is estimated, y^D. The main estimates reported in Table 2 are reported in the central column here. When $y^D = 6.07$, the proportion of non-bunchers α is estimated as the ratio of the counterfactual frequency to the actual frequency in the first bin to the right of the LTU threshold. When $y^D = 6.13$ and $y^D = 6.20$, we use the first two and three bins, respectively.

Table A.9: Robustness Checks for Complementarity Result

	Dependent variable: Average Bunching (b_{av})							
	Empl. below median		Empl. above median		Assets below median		Assets above median	
	OLS	WLS	OLS	WLS	OLS	WLS	OLS	WLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share of Final Consumer Sales	-0.042	-0.029	-0.062	-0.099	-0.096	-0.034	-0.069	-0.035
	(0.041)	(0.017)*	(0.034)*	(0.025)***	(0.042)**	(0.017)**	(0.029)**	(0.020)*
Median Number of Employees					-0.001	-0.000	-0.001	-0.000
					(0.002)	(0.001)	(0.001)	(0.000)
Median Tangible Fixed Assets	-0.018	-0.008	-0.037	-0.002				
	(0.018)	(0.027)	(0.021)	(0.012)				
Constant	0.129	0.105	0.143	0.100	0.184	0.131	0.104	0.106
	(0.021)***	(0.009)***	(0.021)***	(0.011)***	(0.042)***	(0.013)***	(0.018)***	(0.016)***
Observations	16	16	16	16	16	16	16	16
Total obs. in subsample	124,864	124,864	121,122	121,122	142,737	142,737	142,833	142,833

Notes: this table shows robustness checks for the complementarity result, which corresponds to Figure 3 and Table 3 in the main text. Columns 1, 3, 5 and 7 are estimated by OLS, and columns 2, 4, 6 and 8 are estimated by WLS. In the latter, the weights are the inverse of the variance of the bunching estimates. Columns 1-2 use data for firms below the median number of employees (in the overall sample), while columns 3-4 use data for firm above the median number of employees. Similarly, columns 5-6 use data for firms below the median level of tangible fixed assets, while columns 7-8 use data for firms above the median level of tangible fixed assets. The unit of observation is the sector of activity (for details on the definition of the 16 sectors, see section D above). Robust standard errors reported in parentheses. Significance levels: *** = 1%, ** = 5%, and * = 10%.